Gödel's Proof and Wittgenstein's Remarks

Allin Cottrell

November, 2024

1 Introduction

I write as one who has long been fascinated by both "Gödel's proof" (that is, the two theorems in his famous article of 1931) and Wittgenstein's *Remarks on the Foundations of Mathematics* (hereinafter RFM) in which he gave, to all appearances, a dismissive account of Gödel's work.[1] I'm far from being the first to attempt to puzzle out the relationship between the two,[2] but I hope to add something to previous discussion. My text proceeds by means of an iterative exposition of "what Gödel proved", on the one hand, and Wittgenstein's perspective on mathematics, on the other, leading eventually to an examination and assessment of Wittgenstein's remarks on Gödel.

2 What Gödel proved: first pass

Let's start from what a reader of a reasonably serious popular, or semi-popular, account of Gödel's work—for example Nagel and Newman (2001)[3]—might take him to have proved. We'll bypass the many weird and wonderful things that some writers have claimed to derive from Gödel; Franzén (2005) skewers most of these with admirable precision. To circumvent such irrelevancies we must first specify what Gödel is talking about. His topic is a formal system, comprising a set of axioms plus rules of inference, which is capable of representing elementary arithmetic (addition and multiplication) on the natural numbers (non-negative integers). In the first instance he has in mind something like *Principia Mathematica* (Russell and Whitehead, 1910), but his results are not tied to the specifics of PM, and in the following we'll refer to the system in question using the generic label S.

Given all that, in terms of what a reasonably well-informed reader will likely take Gödel to have proved we're looking at one or more of the following statements.

1. He proved that such a system cannot be both consistent and complete.
2. He showed that in any such system there are bound to be propositions that are *true*, but which cannot be proved.
3. He showed that in such a system there must be *undecidable* propositions, which can neither be proved nor disproved.

[1] The proper references: Gödel (1931) and Wittgenstein (1967).

[2] There was no personal relationship; I'm just talking of their work.

[3] Rebecca Goldstein's *Incompleteness* (2005) might seem to belong in this category, but see Feferman's devastating review (2006).

4. He showed that no such system can prove its own consistency.

Most of these statements are correct, provided that the technical terms therein are understood correctly. But statement 2 (which I placed near the top of the list since it's a claim that's often made) needs immediate amendment. Gödel's unprovable proposition—call it G—can be taken as *true* only on condition that the system in question is assumed to be consistent. In other words, he proved that *if* S is consistent, *then* G is true.

OK, so let's work a little on the technical terms, starting with statement 1. Consistency—in its most basic meaning, sometimes called *simple consistency*—is ascribed to a formal system S if and only if there is no proposition p such that S proves both p and its negation $\neg p$ ("not-p").[4] This is a rather weak condition. By analogy with a consistent liar, a consistent formal system may prove all kinds of falsehoods just so long as it doesn't prove both a falsehood and its (true) negation.

A brief digression on this point may be worthwhile. To prove the consistency of a system S it's enough to show that there's at least one well-formed proposition in the language of S that is *not* proved by S. That's sufficient on account of the concept of "material implication" in standard formal logic. Here's how it goes. $A \supset B$ (A implies B) means that you can't have A true and B false. In logical notation, the conjunction that's ruled out is written as $A \wedge \neg B$, '$\wedge$' being the logical "and" symbol. That sounds OK, doesn't it? Then it seems that $A \supset B$ must be equivalent to the negation of that conjunction, namely $\neg(A \wedge \neg B)$. But suppose A is a contradiction, equivalent to $p \wedge \neg p$ for some p. Then A can *never* be true, so $\neg(A \wedge \neg B)$ is guaranteed to be true, regardless of the truth or falsehood of B. So if S proves $p \wedge \neg p$ for some p, any and every B is implied, or in other words is 'proved'. In the ordinary way of things we would not say that "It's raining and it's not raining" implies "London is in Japan", but in standard formal logic this sort of implication is impeccable.

Still dealing with statement 1 above, we now come to the *completeness* of a formal system. This can have more than one sense, but in the context of Gödel's 1931 paper the specific meaning is clear enough. However, to get it right we first need to define an additional term.

A proposition, or well-formed formula (wff), in the language of a formal system S can be proved or disproved (i.e., its negation can be proved)—based only on the axioms and rules of inference of S, without any additional *ad hoc* premises—only if it is a so-called *closed wff* or *sentence*. A sentence is a wff in which any variables that appear are *bound* by quantifiers. Consider this proposition (which is not written in the language of a standard formal system, but could easily be translated into such a language):

$$x + 4 = 7$$

Is it true? Well, it's true if $x = 3$, otherwise false. The variable x is said to be *free*, and the proposition is not of the sort that could be proved or disproved in the manner indicated above. But here are two modified versions:

$$\exists x(x + 4 = 7)$$
$$\forall x(x + 4 = 7)$$

The first variant says that there exists an x (which we take to mean, a natural number) such that $x + 4 = 7$, which is clearly true and ought to be provable in any system designed to encompass basic

[4] The statement that a formal system S 'proves' a proposition p is logician's shorthand. Strictly speaking, what it means is that a *prover* (human or mechanical) can prove p using the axioms and rules of inference of S.

arithmetic. The second says that every x satisfies $x + 4 = 7$; that's clearly false and a formal system of the relevant sort should surely be able to prove its negation. The point is that by *binding* the variable x, via an existential or universal quantifier, we have produced a closed wff that's capable of being true or false in general, provable or disprovable (or perhaps undecidable) as the case may be.

So now we can define completeness for our purposes: a system S is complete if and only if every sentence (closed wff) in its language can be proved (derived as a theorem) or disproved (its negation is a theorem). And Gödel exhibited a sentence—expressible in a large class of formal systems—that (on certain assumptions) can neither be proved nor disproved.

For the moment, let's leave Gödel and develop a few relevant points about Wittgenstein's take on mathematics.

3 Wittgenstein on proof and infinity

A topic on which Wittgenstein lays stress in RFM is mathematical proof. He makes two closely related claims. First, if a purported proof is really to function as such it must be 'surveyable' or 'perspicuous' (terms I take to be more or less synonymous); it must be capable of being 'taken in'. Second, he says that a proof serves to give a definite meaning to the proposition that is proved. If a given mathematical proposition has not been proved, and neither has its negation been proved, its meaning is not fully determinate; its sense remains to some degree 'veiled'. It's the proof that makes clear the meaning of that which has been proved. This sounds quite plausible, but of course is not immune to objection.

Take for example Goldbach's Conjecture—namely, that every even number greater than 2 can be written as the sum of two primes (possibly with repetition). The first few candidates clearly fit the bill—$4 = 2 + 2$, $6 = 3 + 3$, $8 = 3 + 5$, $10 = 3 + 7$ (or $5 + 5$)—and as of this writing mathematicians have searched for counter-examples up to at least $n \leq 4 \times 10^{17}$ without finding any. But to date neither the conjecture nor its negation has been proved. Does the meaning of the conjecture then remain 'veiled'? I suspect that most mathematicians would say that it's already clear enough, but how might Wittgenstein respond?

That question leads us onto another aspect of Wittgenstein's take on mathematics—his skepticism regarding infinity.

There's a distinction between the "potential infinite" and the "actual" or "completed" infinite that goes back to Aristotle. In modern discussions it's sometimes expressed as the difference between a *sequence* $1, 2, 3, \ldots$ that doesn't terminate and an infinite *set* $\{1, 2, 3, \ldots\}$, as in Cantor's set theory. From Aristotle onward, the potential infinite has been generally recognized as an indispensable item in the mathematician's toolkit but the status of the actual infinite has been controversial. Those who reject it typically argue that the Law of Excluded Middle (which asserts that any meaningful statement of purported fact must be either true or false, there being no third alternative) doesn't properly apply in the context of an infinite series. Followers of Cantor, however, are happy to apply the Law to infinite sets.

Wittgenstein was one who vehemently rejected the "actual infinite", and in this respect he was was by no means alone: he was in the company of Henri Poincaré, Leopold Kronecker and Thoralf Skolem, for example.[5] Wittgenstein notes that we're tempted to think of an infinite series as a

[5]Even David Hilbert—who called the transfinite realm opened up by Cantor a 'paradise'—insisted that the proofs

super-long series (a *plenitude*). But that's wrong, he claims: an infinite series is characterized by an *absence*—it doesn't have a last member. And he asks of the infinite case, can we be sure that what we're talking about might not begin to 'flicker' in the 'far distance'.

A simple example of Wittgenstein's skepticism can be found in a question he discussed several time in his writings and lectures on the philosophy of mathematics: Does the sequence 777 occur in the decimal expansion of π? For a Cantorian this question has a definite answer: it either occurs or it doesn't. If it doesn't occur we may never know that, but it's still a fact in good standing, one way or the other. Actually, in relation to Wittgenstein's particular '777' we know the answer today:[6] that sequence first occurs at digit 1589, and it occurs 106 times in the first 100,000 digits—much as we would expect if the digits of π were effectively random, since in that case the probability of that sequence starting at any point in the expansion is 0.1^3 or one in a thousand. But Wittgenstein's question could be restated as: "Does 777 occur in the decimal expansion beyond the number of digits to which π can currently be calculated?" On a probabilistic basis we'd no doubt say that's very likely, but it's not clear that there's any pre-existing fact of the matter.

These points are relevant because Goldbach's Conjecture is a universally quantified claim regarding the infinite series of natural numbers: for *every* natural number x, either x is less than 3, or x is odd, or x can be written as the sum of two primes. Wittgenstein regards as quite comprehensible a claim about the members of a finite series of length n, even if n is very large. But he doubts that claims involving universal quantification over an infinite domain are in good standing.

Two more points are worth considering before we depart from this line of thought. First, what about Fermat's "Last Theorem" (or more properly, "conjecture", since it's very unlikely that Fermat actually had a proof). The proposition here is that the equation

$$x^n + y^n = z^n$$

where n is an integer, has no solution in positive integers (x, y, z) for $n > 2$. Unlike Goldbach's Conjecture, Fermat's has now been proved—to the satisfaction of the mathematical community—by Wiles (1995).[7] So in relation to Wittgenstein's claims about proof we might ask whether Wiles's proof was 'surveyable', and whether it had the effect of clarifying the meaning of Fermat's conjecture. Well, it's clear that the 109 pages of Wiles (1995) cannot be 'taken in' by most of us; nonetheless it seems they *were* surveyable by those with the relevant expertise. That's shown by the fact that Wiles's original attempt at a proof was judged by his peers to be flawed, but then his revised version, following a year of additional work, was seen to be correct. As for fixing the meaning of Fermat's proposal, that may be less clear but at least we can say that Wiles established a linkage with ideas developed in twentieth-century mathematics. As Franzén (2005) notes, Wiles's proof was basically a proof of what's known as the Taniyama–Shimura conjecture for elliptic curves in the semi-stable case, which implies Fermat's claim. Perhaps for a sufficiently adept mathematician that helps in understanding the theorem.

Part of the point of what might seem like a digression on Wittgenstein's ideas about proof is that it prepares the way for this thought: to get a good understanding of what Gödel proved, it will be necessary to take a look at the proof itself. Indeed, Kienzler and Sunday Grève (2016) cite a letter from Wittgenstein to Moritz Schlick in which he says,

of completeness and consistency that he sought must be 'finitary': mathematicians could embrace the transfinite, but could not rely upon it in proving the soundness of their formal apparatus.

[6] See `http://newton.ex.ac.uk/research/qsystems/collabs/pi/`.

[7] See Singh (1997) for an engaging account of Wiles's quest.

> If you hear that someone has proved that there must be unprovable sentences in mathematics, then there is not yet anything astonishing in this, because you have as yet no idea whatsoever what this prose sentence that seems to be so clear is saying. You have, therefore, to go through the proof from A to Z in order to see what it proves.

While I don't intend to go through Gödel's incompleteness proof in anything like full detail, I hope to present enough to "see what it proves". However, before doing so it will be useful to draw out something else from the discussion above of Goldbach's Conjecture and Fermat's Last Theorem.

Fermat's claim has an important similarity with Goldbach, and in fact it's commonly referred to as a *Goldbach-like statement*. Here's what they have in common. First, each one can be put into the form of the statement that every natural number has a certain well-defined property, with the help of disjunction. In Fermat's case: every natural number n is either less than 3 or there are no natural numbers x, y and z such that $x^n + y^n - z^n = 0$. Second, whether a given natural number has the property in question can be determined by means of a definite *algorithm*, a readily programmable test. The property is said to be *computable*. And that means that if the statement is false, it will be possible in principle to find a counter-example; if it's false, it's *decidable*. If it's true, a search for counter-examples would go on forever, and in general we can't know whether it's decidable or not. (Prior to Wiles's proof, many mathematicians thought Fermat's claim was probably true but undecidable.)

Now it's worth noting that the concepts of computability and decidability were developed in the 1930s, prior to the development of what we now know as computers.[8] Turing (1936) developed these concepts by means of what became known as a Turing machine, but this was a thought-object, not bound by requirements of finite memory or finite run-time. So one should bear in mind that if a property is 'computable', in the language of mathematical logic, that by no means guarantees that it can be computed by a real computer on a time-scale of relevance to humans (or even at all).

Now let's get back to Gödel's proof, starting with a key ingredient, his arithmetization of syntax, known as Gödel numbering.

4 Gödel numbering

This is a system which constructs for each distinct well-formed formula in the language L of a formal system S a unique number g (a positive integer). It is designed such that given a valid g (not every positive integer is an instance of g) one can in principle decode this number and retrieve the formula it encodes. Detailed accounts of how this is done, both in Gödel's own system of 1931 and in various alternatives which achieve the same object, are readily available. Here we'll give a brief and incomplete sketch of the method set out by Smith (2007), which is close to Gödel's own; it employs prime numbers in such a way that the possibility of decoding is ensured by the Fundamental Theorem of Arithmetic (every integer greater than 1 is either a prime or has a unique prime factorization).

The fourteen constants in L are assigned the first fourteen odd numbers, and the names of numerical variables $(x, y, z, \dots)$ are assigned successive even numbers.

[8] There were "computers" then, but they were people—mostly women—who performed computations by hand or with the help of mechanical calculators.

sign	0	s	¬	∧	∨	⊃	=	∀	∃	≡	(	)	+	×	x	y	z	⋯
code	1	3	5	7	9	11	13	15	17	19	21	23	25	27	2	4	6	⋯

A formula f including k symbols is given a number equal to the product

$$\prod_{i=1}^{k} \pi_i^{c_i} = \pi_1^{c_1} \times \pi_2^{c_2} \cdots \times \pi_k^{c_k} \tag{1}$$

where π_i is the i^{th} prime by magnitude and c_i is the numeric code of the i^{th} symbol in f. A sequence of m formulae is then assigned the number

$$\prod_{j=1}^{m} \pi_j^{g_j} = \pi_1^{g_1} \times \pi_2^{g_2} \cdots \times \pi_m^{g_m} \tag{2}$$

where g_j is the Gödel number of formula j.

Among the constants are 0, s, $=$ and $+$, with numerical codes 1, 3, 19 and 25 respectively. The symbol s represents the successor operator, which gives the successor of any natural number and can be concatenated to give "the successor of the successor" and so on. The other symbols cited are fairly self-explanatory, representing the number zero, "equals" and addition. In this notation one writes "$2 + 2 = 4$" as

$$\mathsf{ss0} + \mathsf{ss0} = \mathsf{ssss0}$$

Applying eq 1, we find that the Gödel number of this sentence is

$$g_1 = 2^3 \times 3^3 \times 5^1 \times 7^{25} \times 11^3 \times 13^3 \times 17^1 \times 19^{13} \times 23^3 \times 29^3 \times 31^3 \times 37^3 \times 41^1 \approx 2.6 \times 10^{75}$$

This is a seriously big number, 56 orders of magnitude greater than the biggest integer that can be represented precisely in ordinary computer arithmetic ($2^{64} \approx 1.845 \times 10^{19}$, which can accommodate the number of seconds since the Big Bang with ease).

To complete this exercise let's consider a little sequence of formulae, composed of the formula above plus the equally trivial

$$\exists\, \mathsf{y}(\mathsf{s0} + \mathsf{y}) = \mathsf{ss0}$$

(meaning, there exists a number y such that $1 + y = 2$). This second formula has Gödel number $g_2 \approx 3.6 \times 10^{122}$. The Gödel number of the *sequence*, using eq 2, is the integer $2^{g_1} \times 3^{g_2}$. This number cannot be written down in full, nor can it be stored on any computer. What we *can* do is calculate, using floating-point arithmetic, an approximation to the number of binary digits that would be required to store it.

$$B = \log_2(2^{g_1} \times 3^{g_2}) = g_1 \log_2(2) + g_2 \log_2(3) \approx 5.8 \times 10^{122}$$

This B is over 40 orders of magnitude greater than the estimated number of elementary particles in the universe. The term 'Vast' ("very much more than astronomical") coined by Dennett (1995) is apposite here.

It should be clear, then, that nobody has ever actually calculated the Gödel number of a non-trivial formula, let alone carrying out the prime factorization of such a number that's needed to recover the corresponding sequence of symbols in L. Gödel numbering is an in-principle construction in thought, not limited by any consideration of real-world feasibility.[9]

J. Barkley Rosser, who produced a strengthened version of Gödel's proof (Rosser, 1936), also gave a relatively informal exposition of Gödel's own proof in Rosser (1939), and I'll draw on the latter here. Note that Rosser uses L below in the same sense as was given above: to refer to the language of a suitable formal system S.

> When numbers have been assigned to formulas, statements about formulas can be replaced by statements about numbers. That is, if P is a property of formulas, we can find a property of numbers, Q, such that the formula A has the property P if and only if the number of A has the property Q.
>
> Many statements about numbers can be expressed in L, even though all cannot. In particular, if P is properly chosen, we can often express "x has the property Q" in L. If x is taken to be the number of a formula of L, we are then expressing in L a statement about a formula of L. (Rosser, 1939, pp. 55–56)

Rosser goes on to say,

> Gödel chooses for P the property of not being provable in L. So if we denote (as Gödel does) "the formula with the number x is provable in L" by "Bew(x)," then "x has property Q" is equivalent to "not-Bew(x)." (Rosser, 1939, pp. 57–58)

("Bew" is short for the German *Beweisbar*, 'provable'.) In other words Gödel comes up with two predicates that apply to natural numbers, with Bew(x) indicating that x is the number of a sentence that stands at the end of a proof in S, and $\neg$Bew(x) indicating that x is not such a number, so the sentence with number x is not proved by S. It's by no means obvious that it should be possible to define such predicates, but by a lengthy argument involving what Gödel calls *rekursiv* functions (now known as 'primitive recursive' following Kleene (1936)) he shows this is true for a large class of L's.

Then, via what came to be known as the Diagonalization Lemma,[10] Gödel constructs a sentence G which can be read as

$$\text{The sentence with number } g \text{ satisfies } \neg\text{Bew}(g)$$

where it "so happens" that the sentence with number g is G itself.[11] So in effect this sentence asserts its own unprovability. That is its metamathematical reading. As stated above, it's possible in principle to decode g and hence determine the "non-meta" reading of G. This reading—let's call it A—is then a statement of arithmetic that is not provable in S, if G is true.

[9] Gödel's prime-based numbering scheme is completely general but other, more economical, schemes can be employed if there's a known upper limit to the number of symbols in the 'alphabet' of L and the number of individual wffs in a proof. Nonetheless, the integers representing non-trivial formulas and sequences will quickly ascend into 'Vast' territory.

[10] As Smith (2007) notes, Gödel credited Carnap (1934) with the first publication of this Lemma, but it seems that Gödel himself had found it in 1930.

[11] The sense of "diagonal" here is akin to that of the 45-degree line in an x-y graph, where $y = x$.

5 What Gödel proved: second pass

Based on the self-referential sentence, G, that he has painstakingly constructed, Gödel offers two arguments that S is incomplete. These are generally called the *semantic argument*, which is fairly straightforward, and the *syntactic argument* (the 'proof' proper), which takes more work. We'll consider these in turn.

5.1 The semantic argument

We introduced the concept of consistency of a formal system above. The semantic argument relies on a stronger condition on S than consistency, namely that S is *sound*. That means that all the theorems of S are *true*; S may or may not prove all relevant truths but it proves no falsehoods.

Why might anyone feel entitled to make such an assumption? Well, we're talking about a system that employs very standard, uncontroversial, rules of inference which are generally accepted as truth-preserving:[12] they won't lead you from true premises to a false conclusion. If the axioms of S are taken to be true—not merely 'posited'—and the inference rules are truth-preserving, it follows that all theorems will be true. But on what basis can we take the axioms to be true? The basis we need is an *interpretation* of the formalism. In this case the standard interpretation is that the axioms of S are 'about' the natural numbers, and the standard view is that they are indeed true of the natural numbers. These axioms go under the name Peano Arithmetic (PA). Here are the first six of them:[13]

1. $\forall x \neg \mathsf{s}x = 0$
2. $\forall x \forall y (\mathsf{s}x = \mathsf{s}y \supset x = y)$
3. $\forall x (x + 0 = x)$
4. $\forall x \forall y (x + \mathsf{s}y = \mathsf{s}(x + y))$
5. $\forall x (x \times 0 = 0)$
6. $\forall x \forall y (x \times \mathsf{s}y = x \times y + x)$

Given the notation established above, these should be fairly self-explanatory, and moreover clearly true of the natural numbers. For example, axiom 2 just says that if the successor of x equals the successor of y (or in other words $x + 1 = y + 1$) it follows that $x = y$. The relatively complex axiom 4 says in effect that for all x and y, $x + (y + 1) = (x + y) + 1$. Axiom 1 might appear a little mysterious, but it's by way of a stipulation: the natural numbers are just the non-negative integers. This is expressed by the statement that no natural number x has 0 as successor.

Does judging the Peano axioms to be true of the natural numbers imply a platonistic conception of mathematical objects having an objective existence, independent of human thought? I don't think so; if the natural numbers are our own invention (based on how we find nature to behave) we

[12] For example instantiation of universal statements: if we have $\forall x F(x)$ we can conclude that $F(a)$ for any a we like.

[13] Peano Arithmetic is so called after Giuseppe Peano (1858–1932). But note that Peano's original set of axioms differed somewhat from the now-standard listing shown below; Peano assumed a background of intuitive set theory and didn't see a need to define addition or multiplication. See Gillies (1982) for details.

should be in a good position to know what's true of them. We can say, simply, that we recognize these axioms to be a faithful formalization of the idea of natural number that is implicit in everyday arithmetic. For more on platonism see Section 6 below, and for more on the relevance of how nature behaves see Section 7.

The axioms above are supplemented by a schema for mathematical induction: if 0 has property P, and the successor of x has property P whenever x has property P, then for all x, x has the property P, or

$$P(0) \wedge (P(n) \supset P(\mathsf{s}n)) \supset \forall x\, P(x)$$

In principle we could make a single axiom of this, if we could preface it by $\forall P$ ("for all properties P"), but that can't be done in the sort of system under discussion. S is a *first-order* system, in which we can quantify over variables ranging over the natural numbers but we can't quantify over properties. That's why the above is referred to as a *schema*; in effect we need a distinct axiom that instantiates the schema for each case where we appeal to induction. The induction schema is the only component of PA about which anyone might have defensible qualms, it seems to me, but even it is a very standard component of mathematical reasoning, in the context of definite formal systems or otherwise. Wittgenstein expressed puzzlement over mathematical induction in his 1939 lectures on the Foundations of Mathematics (Diamond, 1976) but he did not pronounce the method illegitimate. That would have flown in the face of his avowed intent not to tell mathematicians how to do mathematics based on philosophical considerations; he wished to limit his criticisms to the questionable philosophical gloss that mathematicians are sometimes tempted to give their results, as in what he called "puffed-up proofs".

So for the moment let's assume that S is indeed sound. We now wish to show that S does not prove G, and neither does it prove $\neg G$ (or in other words, G is undecidable in S).

Suppose S *does* prove G; then it would prove a falsehood, since G states that G is *not* provable. But if S can't prove falsehoods it can't prove G. However, if G is not provable then it's *true*, because that's just what it says. And now, if we take the truth of G to be established, it follows on the same semantic grounds that $\neg G$ can't be proved either since $\neg G$ must be false and, once again, S doesn't prove any false statements.

Before leaving the semantic argument, it's worth raising the question, what about the arithmetical reading of G? That too—we called it A above—should be true but unprovable in S, but what might it look like? Since computing the actual Gödel number, g, of G is infeasible in practice, as is the prime factorization of g that could in principle lead us to an explicit statement of A, it might seem that A must remain inscrutable. Something of the sort is suggested by (Goldstein, 2005, p. 167): "it's going to be a weird arithmetical proposition, what with all the fiddlefaddling that gets us to it." Similarly, (Steiner, 2001, p. 272) says that the Gödel sentence "has no mathematical interest: the sentence, regarded as a number-theoretic one, is so long as to be unsurveyable."

Surprisingly enough, A turns out not to be quite so weird and inscrutable. That's because G amounts to a generalization regarding a property of the natural numbers—namely, every natural number has the property of not being the Gödel number of a proof of G—and "not being the Gödel number of a proof of G" turns out to be a primitive recursive property. This places A in the category of *Goldbach-like* statements, which we came across in Section 3.

Stephen Kleene gives an account of developments following Gödel's presentation at the Königsberg Conference of 1930 (for more on which see Section 6).

> At that time Gödel only had undecideable propositions which were finitary combinatorial in nature, and von Neumann asked whether number-theoretic undecideable propositions could also be constructed. Gödel replied that they would have to contain concepts quite different from those occurring in number theory like addition and multiplication. Gödel was astonished when slightly afterward he succeeded in turning the undecideable proposition into a polynomial equation preceded by quantifiers over the natural numbers (Theorem VIII). (Kleene, 1986, p. 137)

In his "Postscriptum" of 1964 to the printing of his 1934 lectures on undecidable propositions, Gödel was able to go further.

> By slightly strengthening the methods used above..., it can easily be accomplished that the prefix of the undecidable proposition consists of only one block of universal quantifiers followed by one block of existential quantifiers, and that, moreover, the degree of the polynomial is 4. (Gödel, 1986b, pp. 370–371)

The "above" here refers to section 8 of his text, titled "Diophantine equivalents of undecidable propositions". A Diophantine equation is a polynomial equation with integer coefficients; we've already seen an example in Section 3, namely Fermat's Last Theorem. Gödel is in effect equating his undecidable proposition to one regarding the solution of a certain Diophantine equation, which can be put into the canonical form of

$$(Q)[P = 0]$$

where (Q) denotes a sequence of quantifiers and P denotes a polynomial with natural-number coefficients.

That this fact is no mere curiosity was shown when Yuri Matiyasevich (1970) completed work begun by Julia Robinson, Martin Davis and Hilary Putnam. This work, spanning over twenty years, produced a result known as the MRDP theorem, which says that there exists no general algorithm to decide whether a given Diophantine equation has a solution among the integers (Davis, 1973). This amounts to an incompleteness proof that stands alongside Gödel's, but does not depend on Gödel's apparatus of the self-referential sentence.

5.2 The syntactic argument

Gödel had no doubt that the system he was considering was sound, but he was well aware that the semantic argument, which rests on the assumption of soundness, would not satisfy the formalists. Hilbert insisted that validation of a formal system should proceed in purely syntactical terms, which requires that the axioms be deinterpreted and talk of 'truth' be eschewed. In the main section of his 1931 paper, Gödel therefore fell back on the assumption of consistency; this property can be defined in syntactical terms, and as noted above it is a weaker condition than soundness.

The first point to notice is that if the system S is consistent, it cannot prove its Gödel sentence G. That's because G "says that" G is *not* provable, so a proof of G in S would amount to a proof of G's contrary, $\neg G$. That would yield a contradiction, which cannot happen in a consistent system.

This point can be unpacked a little and made somewhat more precise. If G were a theorem of S this fact would itself be provable, via the arithmetization of syntax: the Gödel number of G would have the computable property of being the number of a formula that stands at the end of a valid proof in S, something that can be verified within S. But G is a provable fixed point ('diagonalization') of the property of *not* being a theorem of S, meaning that a proof of G in S would also prove the negation of G, so S would be inconsistent, contrary to assumption.

To complete his argument that G is undecideable in S, Gödel also has to show that S cannot prove $\neg G$. That turned out to require a stronger condition on S than simple consistency, a condition that he dubbed ω-consistency.

Let the expression "$\varphi(x)$" mean that some predicate φ is satisfied by a natural number x. (Simple examples include "x is divisible by 2", "x is prime", and so on.) Now suppose that for each natural number, x, taken in turn, S can prove that $\varphi(x)$ holds. So S can prove $\varphi(0)$, $\varphi(1)$, $\varphi(2)$, ... One might then expect that S should also be able to prove the generalization, $\forall x\, \varphi(x)$, but that's not necessarily the case. If the generalization cannot be proved that's a case of ω-incompleteness, which Peter Smith (2007) describes as a "regrettable weakness" for a formal system. But ω-inconsistency is worse: it means that despite proving each individual case, the system proves the *negation* of the generalization, $\neg\forall x\varphi(x)$. Conversely, an ω-inconsistent system might prove the generalization $\forall x\, \varphi(x)$ but also prove that there exists an exception, $\exists x\, \neg\varphi(x)$. In neither case is this a flat-out proof of $p \wedge \neg p$ but it certainly looks like a contradiction ("equally intolerable", says Quine (1953)). It's easily shown that if S is ω-consistent—that is, it is not ω-inconsistent—it necessarily satisfies simple consistency. But the converse does not hold, so ω-inconsistency is the stronger condition.

Now let's define $\varphi(x)$ as the predicate Gödel is interested in, to complete his proof of the undecidability of G in S—namely, x is *not* the Gödel number of a proof of $\neg G$ in S. Suppose Gödel is able to prove $\forall x\, \varphi(x)$. The trouble is that if S is ω-inconsistent this does not rule out the possibility of a proof of $\exists x\, \neg\varphi(x)$, which is to say that there exists some (unspecified) x that *is* such a Gödel number. Hence, as stated above, Gödel needs to assume the ω-consistency of S to clinch his argument.

However, it later transpired that the stronger assumption was not strictly required: Rosser (1936) produced a version of Gödel's undecidability proof that depends only on simple consistency of S. The counterpart of Gödel's sentence G is a somewhat more complex Rosser sentence R which says, in effect, "If a proof of the sentence with Gödel number r exists in S, there also exists in S a shorter proof of the negation of that sentence." And of course the sentence with Gödel number r turns out to be R itself. For discussion of Rosser's result as a strengthening of Gödel's see Kleene (1952), Mendelson (1964), Smith (2007).

5.3 The unprovability of consistency

The last point alluded to in Section 2 concerned Gödel's proof that a formal system of the sort under discussion cannot prove its own consistency—hence dashing one of Hilbert's cherished hopes. We're now in a position to address this in a little more detail.

As noted above, to show that a system S is consistent it's sufficient to exhibit a proposition in S that cannot be proved (since if S were inconsistent any arbitrary proposition could be proved). Now of course the Gödel sentence G "says of itself" that it's not provable in S. And as we have seen, on the assumption that S is consistent G must be true: consistency of S implies G. We're now noting in addition that the truth of G (that is, its unprovability) implies the consistency of S. This is a case of bi-implication or logical identity. But then the impossibility of proving that G is true within S leads directly to the impossibility of proving the consistency of S within S.[14]

This is the merest sketch of an argument, and although Gödel's own version in 1931 was more extended, Franzén (2005) judges that it "consisted mostly of handwaving". Full formal details were eventually provided by Hilbert and Bernays (1939), Hilbert (or at least Bernays, who was by all accounts the main author of this work) having meanwhile come to terms with the fallout from Gödel's work.

5.4 Can incompleteness be remedied?

A further question arises from Gödel's two proofs: does his perspective offer any means of "fixing" the incompleteness of S? Well, only at the price of an infinite recursion—which we might take as meaning, not really. If the Gödel sentence, G, of the system S is taken to be true, we could define a system S^+, identical to S except that the original G is added as an axiom. If, as per the second proof, G is logically equivalent to $\mathrm{Cons}(S)$, this would amount to making the consistency of S an axiom of S^+. The trouble is that a new Gödel sentence could then be found for S^+, which could in turn be added as an axiom for a further system S^{++}. And so on, *ad infinitum*.[15]

5.5 Gödel's reach

There's something rather surprising about the reach of Gödel's proofs. On the one hand, it was quite quickly realized that his results were not specific to the system of *Principia Mathematica*, or to particularly close relatives of PM, but applied to any formal system capable of representing elementary arithmetic (addition and multiplication) on the natural numbers. Gödel had originally thought he had more work to do to demonstrate this generality, which is why he appended "I" to the title of his 1931 paper, but it soon became clear that this would not be necessary. On the other hand, we find that systems that formalize arithmetic on the *real* numbers are not prey to Gödelian incompleteness Tarski (1951). And neither are systems that confine themselves to either addition (Presburger, 1929) or multiplication (Skolem, 1931) of the natural numbers. It's the particular combination of addition and multiplication, confined to the natural numbers, that turns out to resist completability.

6 The platonism connection

It's time to say something about platonism. This is relevant to our topic because Gödel was a convinced platonist and Wittgenstein a convinced anti-platonist, which might lead us to expect them to hold opposed views on specific points such as the meaning of Gödel's proof, possibly compounded by a degree of mutual incomprehension.

In the philosophy of mathematics, platonism is the idea that (at least some) mathematical objects exist independently of human cognition and practices, in an immaterial realm of their own. Humans are able to apprehend the properties of these objects (more or less clearly, more or less accurately) via the faculty of reason, and the task of mathematics is to discover, as best we can, such properties.

[14] A prior condition is needed if an impossibility proof along Gödel's lines is to go forward: the proposition $\mathrm{Cons}(S)$, stating that S is consistent, must be *expressible* in the language of S. That can't be taken for granted, but it turns out to be so; $\mathrm{Cons}(S)$ is primitive recursive.

[15] In his PhD thesis, under the supervision of Alonzo Church, Alan Turing (1939) seemed to be willing to explore this route, to the extent of transfinite recursion, though not without some reservations. This reader finds it difficult to determine how seriously Turing regarded such a project.

Platonism is a matter of degree. One might, for example, grant a platonic existence to the natural numbers but reckon that more complex mathematical objects are our own constructions—there's the well known statement ascribed to Kronecker, "God created the integers, all else is the work of man." A hard-core platonist, however, may extend the courtesy much further, positing a platonic existence for such objects as Cantor's transfinite sets.

It seems that many working mathematicians subscribe to platonistic views to some degree,[16] but among philosophers platonism is far from fashionable. *Naturalism* is a general tendency of modern philosophy, and it seems clear that independently existing mathematical objects would have to be supernatural.

I said above that Gödel was a convinced platonist.[17] A startling index of this is his remark, "It might be that there are only finitely many integers" (Wang, 1996, p. 302) The context here is Gödel's parallel adherence to *fallibilism*, the idea that we can never have absolutely certain knowledge of anything. It may seem evident to us that the sequence of natural numbers has no end, but what seems evident to us may not be true. There is—'out there', so to speak—a *fact of the matter* as to whether the (independently existing) natural numbers go on forever or not. Gödel even gives an illustration: for all we know, 10^{10} could be the greatest natural number.

Gödel's platonism extended to the Cantorian transfinite. The so-called Continuum Hypothesis (CH) is a supposition made, and considered of great importance, by Cantor. (It concerns the relative size of various infinite sets; I will not attempt to explain it here.) To date nobody has been able to prove or refute CH. It has been shown to be *independent* of the now-standard version of set theory—Zermelo–Fraenkel plus the Axiom of Choice, or ZFC—in the sense that neither CH nor its negation is inconsistent with the axioms of ZFC. Gödel devoted considerable effort to obtaining a proof or refutation of CH. He did not succeed, but he was motivated by the belief that CH *must* be either true or false, as a matter of platonic fact. Anti-platonist logicians will likely suspect that the Cantorian concepts at play in CH are just not sufficiently well-defined to make the truth or falsity of the hypothesis an answerable question (Feferman, 1998), so that the question loses its interest.

In discussion with Hao Wang during his time at the Institute for Advanced Studies, Gödel said that his platonistic, or *objectivistic*, convictions dated to well before his famous paper of 1931, and indeed before his first notable publication, a proof of the completeness of first-order logic (Gödel, 1930). Wang asked Gödel why he thought there had been such a long pause between Skolem's work in the early 1920s and his own 1930 completeness result, for which (at least with hindsight) Skolem had laid the foundation. Gödel's reply is interesting: he ascribes the delay to anti-platonistic thinking.

> This blindness (or prejudice, or whatever you may call it) of logicians is indeed surprising. But I think the explanation is not hard to find. It lies in a widespread lack, at that time, of the required epistemological attitude toward metamathematics and toward non-finitary reasoning. (Wang, 1996, p. 204)

In similar comments, Gödel says that his "objectivistic conception of mathematics and metamathematics in general, and of transfinite reasoning in particular, was fundamental" to his work in logic (Wang, 1996, p. 73); and he says of Hilbert's "On the infinite" (Hilbert, 1926): "Here again the anti-Platonistic view was hampering mathematics" (Wang, 1996, p. 85).

[16] Hacking (2014) offers an interesting discussion of this point—see chapter 6 in particular.

[17] As a general statement that's uncontroversial, but there is some room for doubt as to when he acquired that conviction, a point to which we'll return briefly below.

Feferman (1998) finds reason to doubt Gödel's retrospective account of his own early views. The first occasion on which Gödel publicly indicated his adherence to platonism was in a discussion of Bertrand Russell's philosophy (Gödel, 1944). Prior to that, in a 1933 lecture (Gödel, 1995), he had said of his account of foundational issues up to a certain point, "The result of the preceding discussion is that our axioms, if interpreted as meaningful statements, necessarily presuppose a kind of Platonism, which cannot satisfy any critical mind and which does not even produce the conviction that they are consistent." It is plausible that when Gödel—who was anyway generally reticent—was in a relatively junior position he might have felt disinclined to advertise his espousal of an unfashionable philosophy. But by 1933, when he was widely recognized as a leading figure in his field, is it plausible that he would willingly traduce his own philosophical commitments in a public lecture?

Anyway—leaving this historical puzzle aside as probably insoluble—the more interesting question is whether Gödel's key results of 1931 somehow depended on platonism. A weak version of this idea might run along these lines: Gödel's (presumed) platonism motivated him more strongly than others to believe that definite results ought to be available in certain difficult areas. A strong version might say that Gödel's results were somehow *logically* dependent on platonistic premises. There may be something to be said for the weak version but the strong version seems wrong. Plenty of non-platonist (and outright anti-platonist) logicians[18] have recognized Gödel's 1931 argument as compelling—at least the syntactic variant if not also the semantic one.

It may be relevant to note the initial reaction to Gödel's incompleteness result. He gave a preview of his findings at a conference on 'The epistemology of the exact sciences' in Königsberg in September of 1930. Several luminaries were present to expound various views on the foundations of mathematics: Rudolf Carnap to represent Bertrand Russell's logicism; John von Neumann for Hilbert's formalism; Arend Heyting for Brouwer's intuitionism; and Friedrich Waismann for Wittgenstein's views. Gödel's presentation on incompleteness was relegated to a session on the last day when some attendees were already leaving. At this point just one of those present understood what Gödel was saying, and grasped its importance: not surprisingly it was von Neumann, not a noted platonist.[19]

That said, it's also true that Gödel's "true but unprovable" sentence was quite widely taken to validate platonistic thinking—Shanker (1988) provides historical evidence to that effect. Gödel's distinguishing mathematical truth from provability was open to an appropriation in which 'truth' was a matter of conformity with a platonic reality.

I'll touch on an extension of this point. It has seemed to some—notably J. R. Lucas (1961)—that Gödel's incompleteness proof can be parlayed into a proof that the human mind surpasses any machine in its ability to recognize mathematical truth. The gist of the argument is that machines are limited to grinding out theorems from axioms plus rules of inference while humans are able to 'see' that Gödel's unprovable sentence is true. There's a lot wrong with this, and Franzén (2005, chapter 6) offers a scrupulous takedown; I'll confine myself to a few remarks. Insofar as humans can 'see' the truth of a Gödel sentence, G, at all, this is to do with its metamathematical reading, and relies on the argument that if S is consistent then G must be true (whatever its arithmetical reading might be). That rule can easily be built into a machine. Although Gödel was able to tell us

[18] Among whom we find Solomon Feferman, the principal editor of Gödel's *Collected Works*.

[19] Keynes once described Piero Sraffa as one "from whom nothing is hid" (Keynes, 1933). John von Neumann was another such. He noticed the corollary, not mentioned in Gödel's 1930 talk, that the consistency of a formal system of the sort that Gödel described could not be proved within the system itself, and remarked on this in a letter to Gödel, who had in fact already arrived at that conclusion.

a surprising amount about the arithmetical reading, A, of his unprovable sentence, nobody has ever seen A in full and the idea that a human might 'see' it to be true is absurd. No mathematician has claimed to 'see' the truth or falsity of much simpler arithmetical propositions such as Goldbach's Conjecture or Fermat's Last Theorem. Moreover, if anyone had made such a claim it would have cut no ice; the response of the mathematical community would have been, "So you say. Let's see a proof."

7 Wittgenstein on mathematics: the basics

We have already noted that Wittgenstein was an anti-platonist, surely an uncontroversial statement. The position in the philosophy of mathematics that is diametrically opposed to platonism goes under various names: *conventionalism*, suggesting that what are sometimes taken to be 'truths' about mathematical objects are really propositions that have been made unassailable by their adoption as conventions; or *constructivism*, suggesting in a similar way that mathematical 'truths' are human constructs (although this label may have other connotations). One of the schools of thought mentioned above in the context of the Königsberg conference, namely Brouwer's intuitionism, might also be claimed as the opposite of platonism. So was the Wittgenstein of RFM a conventionalist, a constructivist or an intuitionist? None of the above, really, although he was sympathetic with Brouwer's point of view in certain respects.

Michael Dummett, in a well known paper (Dummett, 1959), described Wittgenstein as a "full-blooded conventionalist". You can certainly find some statements in Wittgenstein's work that seem to justify that label, for example, that the mathematician is "an inventor, not a discoverer." But I agree with Gerrard (1996) that this is a superficial judgment: taking Wittgenstein's work on mathematics as a whole, he is trying to thread the needle between platonism and conventionalism—as Gerrard puts it, "a philosophy of mathematics between two camps."

One message that comes across loud and clear from RFM is that Bertrand Russell's project of providing a logical foundation for mathematics (*logicism*) is misguided. Russell's logical apparatus is a calculus in its own right, says Wittgenstein—possibly useful in some ways, but in no way prior to or 'foundational for' ordinary arithmetic.

In my reading, Wittgenstein's point is that ordinary arithmetic already has a perfectly good 'foundation' in our practices of counting and measurement. And these practices are useful to us—have become part of our way of life—due to some very general facts of nature that we're inclined to take for granted but that in principle *could* have been otherwise. There's a multitude of objects of interest to us in our surroundings that do not generally fuse or divide spontaneously, so that counting "works" quite stably most of the time. Similarly, measurement of magnitudes in which we're interested turns out to be mostly stable. Some things we want to measure can change in size with temperature—as can, for example, metal rulers—but these changes turn out to be fairly predictable and do not vitiate our intent to measure. If we get a notion of "natural number" via counting, we can get a notion of "real number" via measurements of progressively greater accuracy.

Here's my main text from RFM (Wittgenstein, 1967, p. 14):

> Put two apples on a bare table, see that no one comes near them and nothing shakes the table; now put another two apples on the table; now count the apples that are there. You have made an experiment; the result of the counting is probably 4. (We should present the result like this: when, in such-and-such circumstances, one puts first

> 2 apples and then another 2 on a table, mostly none disappear and none get added.) And analogous experiments can be carried out, with the same result, with all kinds of solid bodies.—This is how our children learn sums; for one makes them put down three beans and then another three beans and then count what is there. If the result at one time were 5, at another 7 (say because, *as we should now say*, one sometimes got added, and one sometimes vanished of itself), then the first thing we said would be that beans were no good for teaching sums. But if the same things happened with sticks, fingers, lines and most other things, that would the end of all sums.
>
> "But shouldn't we then still have $2 + 2 = 4$?"—This sentence would have become unusable.

As Wittgenstein indicates, addition arises naturally from counting. He goes on to show how multiplication can be based on iteration of addition, and exponentiation on iteration of multiplication—all meaningful for us so long as the world continues to cooperate, for the most part!

At one point in his lectures on the Foundations of Mathematics, in Cambridge in 1939, Wittgenstein cited Russell:

> Russell said, "It is possible that we have always made a mistake in saying $12 \times 12 = 144$." But what would it be like to make a mistake? Would we not say, "This is what we do when we perform the process which we call 'multiplication'. 144 is what we call 'the right result' "? (Diamond, 1976, p. 97)

If Russell said such a thing it seems likely it was somewhat tongue-in-cheek, but perhaps designed to suggest that the rigorous logical apparatus of *Principia Mathematica* could *in principle* correct a long-standing error in received school-room arithmetic. But Wittgenstein's dismissal is surely right: if a purported proof in PM showed 12×12 to be anything other than 144, we would conclude that there was an error in the proof, or in the system of PM itself. The received result is much too thoroughly grounded in our practice of counting to be given up on the say-so of a system such as Russell's.

Returning to the theme of pigeon-holing Wittgenstein's views, I said above that he was sympathetic with Brouwer's point of view in certain respects (their approval of finitism would be one such point of contact). But here's one respect in which the two diverge. Brouwer liked to talk of mathematics as a "free creation" of the human mind. I don't detect anything of the sort in Wittgenstein. Arithmetic is not of itself an empirical science, but its rules are an abstraction from what we commonly observe, and therefore come to treat as paradigmatic—so, according to Wittgenstein, neither a "free creation" nor something based on insight into a platonic realm.

8 Wittgenstein on Gödel

We are now, hopefully, in a position to assess Wittgenstein's remarks on Gödel in *Remarks on the Foundations of Mathematics*. But before engaging with the text it may be worth offering a brief account of reactions to these remarks to date. As I see it, they have unfolded in three main phases.

8.1 Reactions to Wittgenstein's remarks

Phase one. Following the first publication of RFM in 1956, five years after Wittgenstein's death, several logicians reviewed the book, notably Bernays (1959) (Paul Bernays was Hilbert's student, his assistant, then his collaborator); Goodstein (1957) (R. L. Goodstein was a student of Wittgenstein's); and Kreisel (1958) (Georg Kreisel was also a student of Wittgenstein's, and one who engaged in personal discussions with Wittgenstein on mathematics following the Second World War).[20] All of these authors found something to agree with in Wittgenstein's philosophy of mathematics. For example, both Bernays and Kreisel explicitly endorsed his view that Russell's logicism ultimately failed, in that it could not provide a firmer foundation for arithmetic than the counting that Wittgenstein had cited; indeed, a Russellian proof of $2 + 2 = 4$ was in the last analysis parasitic on counting. But their reviews were quite scathing on the topic we're concerned with here. The gist: Wittgenstein had not really understood Gödel's work, and his comments were worth little. Bernays: "his discussion of Gödel's theorem ... suffers from the defect that Gödel's quite explicit premiss concerning the consistency of the formal system under consideration is ignored." Goodstein: "unimportant and throws no light on Gödel's work". Kreisel: "Wittgenstein's views on mathematical logic are not worth much because he knew very little and what he knew was confined to the Frege–Russell line of goods."

Phase two. Some time later several "revisionist" authors published emphatic defences of Wittgenstein. Wittgenstein had not misunderstood Gödel, they claimed; rather the logicians had failed to grasp Wittgenstein's point. Examples include Shanker (1987), Shanker (1988), Floyd (1995, 2001) and Floyd and Putnam (2000). Meanwhile Berto (1996), although not explicitly endorsing Wittgenstein's remarks, argued that they may be of interest in the light of relatively recent developments such as Paraconsistent Logic.[21]

Phase three. The debate continues. Steiner (2001) and Bays (2004), for example, have contested the revisionist reading, while Rodych (2002), for example, has supported it.

We'll return to some of these contributions below, but for now: to business with the text.

8.2 Wittgenstein's text

The most substantive comments on Gödel are to be found in Appendix I of RFM on pages 50–53 (dated to 1937 by the editors).[22] I'll examine this text, quoting extensively but interspersing some comments.

> 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?
>
> 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 a proposition used in the language game? And the 'assertion

[20]According to Monk (1990, p. 498) Kreisel at age twenty-one won Wittgenstein's approbation as "the most able philosopher he had ever met who was also a mathematician".

[21]For an account of which, see https://plato.stanford.edu/entries/logic-paraconsistent/.

[22]The exact same text appears in the 1978 revised edition of RFM, but as Appendix III, with different page numbering.

> 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.

I think one can accept Wittgenstein's equation of asserting that p is true with simply asserting p, without prejudice to what follows. Continuing with §6:

> 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.

That is true of "Russell's game" as understood by Russell himself, but Wittgenstein is here implicitly excluding the possibility that Gödel might have found a third ground for asserting a proposition expressible in Russell's system. I'll come back to this point below.

> 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?

This I don't really get. As a proposition of physics, take for example Newton's $F = ma$. In Newton's theory we'd think of all the terms in this equation as real-valued variables. I suppose that could be specialized to the natural numbers for the sake of argument, but how would you write Newton's proposition in the notation of a formal system that supports elementary arithmetic? One might try

$$\forall m \forall a \exists f (f = m \times a)$$

(For all natural numbers m and a there exists a natural number f such that f equals m times a.) That would itself be *provable* in a system such as Russell's but it has nothing to do with Newton, and I don't see how you could make it even *appear* to have anything to do with Newton given the resources of the sort of system in question. Let's continue with §7.

> 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 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*.

Again Wittgenstein is ruling out—*a priori*, one might say—the possibility that there could be grounds for saying that an unprovable proposition p in a formal system S is true, other than by interpreting p as belonging to a different language game—that is, via a sort of pun. But might not Gödel have come up with something that is not so much a "different language game" or "another system" but an *extension* of a Russell-type system—and an extension which is not easily dismissed as "foreign" and hence inadmissible?

Wittgenstein's next numbered point:

> 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 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."

Leaving aside the business of portraying of Gödel as a *naif* seeking advice on what to make of his theorem, this is not totally out of order as a highly compressed account of his argument. But Wittgenstein aims to problematize "suppose it were false", as we find on page 51.

> Just as we ask, " 'provable' in what system?", so we must also 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.

Here it seems that Wittgenstein does not take on board that "the interpretation that [the relevant Gödel sentence] is unprovable" is not optional for Gödel; it's not something that he could decide to "give up" since it's something he had ensured with a rigor that had, by 1937, been fully recognized by the community of mathematical logicians. By the contrapositive, Gödel's "suppose it is false" does *not* mean supposing that its negation has been proved in the formal system in question (but simply that it's false). Another perspective on this point is that Wittgenstein is not allowing Gödel's assumption that the system in question is ω-consistent, in which case it's provably impossible that the negation of P could be proved. There's further discussion of this point in Section 8.4.

Let's proceed to §9 and §10.

> 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.
>
> 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*"?

Again we have the theme of giving up an interpretation. It seems that Wittgenstein wants to drive a wedge between whatever it is that Gödel has definitely proved in mathematical terms and the natural-language interpretation of the Gödel sentence P as saying "P is not provable"—between the ‘proof’ and the ‘prose’, as Floyd (2001) puts it. When the ‘prose’ is given in the simple (and in fact, imprecise) form just stated, it seems quite plausible that there *could* be slippage here, that the interpretation might be put in question without impugning any of Gödel’s technical work. However, the plausibility of slippage is diminished if the prose is put into a more precise form. Here I’ll revert to use of the generic S to indicate the formal system in question in place of “Russell’s system”. Given that every wff of S has a Gödel number, and that g is the Gödel number of P, one might try the following.

> *Prose 2*: There exists no integer x such that x has the property ψ of being the Gödel number of a proof of the sentence with Gödel number g—where ψ is a primitive recursive property, defined in syntactical terms and indicating that the sentence with Gödel number g can be reached from the axioms of S by application of the rules of inference of S.

That statement could no doubt be improved upon; my point is that something like it can serve as a bridge, being relatively close to the formal argument but also pretty clearly sustaining the conclusion that P is not provable in S. The question would then be: Now where are you going to insert the wedge? If I really believed I had constructed a proof of P, I would have to conclude that *Prose 2* was incorrect, and that would not just be a matter of interpretation, it would have to involve a formal error.

Next Wittgenstein tries out the idea that the unprovability of P might be proved in S:

> 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?

Gödel’s position is clearly that there cannot be a contradiction here. Since he’s assuming that the system S he’s working with is consistent, there can’t be a proof in S of the unprovability of P. On that view, Wittgenstein’s “Let us suppose. . . ” is illegitimate (other than as the starting point of a possible *reductio*), and the question of whether the contradiction does any harm simply doesn’t arise.

> 12. Is there 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.

Wittgenstein’s views on the relative unimportance of contradiction, and the “superstitious fear and awe of mathematicians in face of contradiction” (RFM, p. 53), are of interest in their own right. But they’re not strictly relevant here because Gödel’s sentence does not have the same status as the Liar paradox of Epimenides. “I am lying” does indeed lead us around a useless circle, as Wittgenstein says. But “this statement is not provable” (a rough translation of Gödel’s sentence) does not imply "this statement is provable"; there's no circle, no paradox. It may be said that Gödel himself muddied the waters by talking of an "analogy" between his argument and the Liar paradox in the introduction of his 1931 paper (Gödel, 1986a, p. 149) but on reading the paper as whole[23] it's clear that his self-referential sentence is neither intended as paradoxical nor in fact paradoxical.[24]

§13 just being an aside (a rumination on 12) I'll proceed to the next point, on p. 52.

> 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 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.

Bernays (1959, p. 23) jumps on the last statement above: "As a matter of fact, such impossibility proofs usually proceed via the derivation of a contradiction." However, Steiner (2001, p. 272) suggests that what Wittgenstein meant by "a contradiction" here was not a *reductio* but a *paradox*. That would make his statement fair comment, had Gödel's argument been built on a paradox as Wittgenstein seems to have believed. Actually Gödel's second proof, in particular, offers a nice example of just the effect that Wittgenstein treats as paradigmatic: people *did* in fact give up trying to find a consistency proof for S within S when Gödel showed it couldn't be done.[25]

Wittgenstein's §15 and §16 return to his claim that a proof (alone) makes clear *what* is proved. If "P is unprovable" were proved, he says, that would show us the sense of the unprovability proposition. But "if it is unproved then *what* is to count as a criterion of its truth is not yet *clear*, and—we may say—its sense is still veiled." To repeat what was said earlier, Gödel's proof of the unprovability of the sentence G (P in Wittgenstein's notation) is a *reductio*, which depends on the assumption that the system S is consistent and thereby rules out the possibility of a direct proof (which would *ipso facto* prove a contradiction).

Elsewhere in RFM Wittgenstein expresses some worries about proof by *reductio* but seems in the end to come to terms with the method. In Part IV (dated to 1942–3 by the editors) the worry is put in this way:

> The difficulty which is felt in connexion with with *reductio ad absurdum* in mathematics is this: what goes on in this proof? Something mathematically absurd, and hence unmathematical? How—one would like to ask—can one so much as assume the math-

[23] Had Wittgenstein read the whole paper? Kreisel (1998, p. 119) speaks of an occasion in the 1940s when Wittgenstein wanted Kreisel to "tell him Gödel's proof"—which, Kreisel says, "Wittgenstein had never read, having been put off by the introduction."

[24] The title essay in Quine (1976) gives a broad definition of paradox, including what he calls "veridical paradox", i.e. an argument with a conclusion which at first seems impossible but nonetheless turns out to be correct. In that sense only can Gödel's incompleteness proof be termed a paradox.

[25] Nobody was forced to give up looking for a proof of the arithmetical Gödel sentence. Since nobody knew exactly what it said, nobody was looking for a proof.

> ematically absurd at all? That I can assume what is physically false and reduce it *ad absurdum* gives me no difficulty. But how to think the—so to speak—unthinkable?!

But this is immediately followed by what looks like a resolution:

> What an indirect proof says, however, is: "If you want *this* then you cannot assume *that*: for only the opposite of what you do not want to abandon would be combinable with *that*."

Very well; read "S is consistent" for *this* (which Gödel does not want to abandon) and "G is provable" for *that* and you have the gist of Gödel's argument.

Wittgenstein's §17 and §18 attempt to explore the possibility that either G or its negation might be proved directly (i.e. as theorems of "Russell's system"). Since both of these cases are expressly ruled out by Gödel's formal argument I proceed to §19 (p. 53).

> 19. You say: "..., so P is true and unprovable". That presumably means: "Therefore P". That 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 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 this assertion plausible to me, since you can make no use of it except to do these bits of legerdemain?

This resembles a comment we've seen made by others, regarding the supposedly outlandish quality of the arithmetical reading, A, of the Gödel sentence—the châlet on Mount Everest—although in Wittgenstein's case it is, so to speak, weaponized. However, as noted above, Gödel quickly developed a good general idea of what his sentence could look like, namely a Diophantine equation. For sure, not one that would ever have been the subject of a conjecture by any human mathematician, but also not an unrecognizable sort of proposition. Anyway, the purpose of Gödel's assertion was not to respond to interest in A as such but to argue that a consistent S could not be complete.

8.3 Reflections on the remarks

It will by now be evident that I do not find Wittgenstein's remarks on Gödel's (first) proof to be apt, either as a critique or a clarification. This is disappointing for one such as myself who sees a great deal of value in other aspects of Wittgenstein's philosophy of mathematics. I did not embark on this project with a *parti pris* against Wittgenstein—far from it. I wished to attain a better understanding of what Gödel proved, and hoped that Wittgenstein's remarks might be illuminating, but came to the conclusion that they were not.

I think I know "where Wittgenstein was coming from": he was motivated by his antipathy to platonism, to which he saw Gödel's work as an on-ramp. I agree with Mark Steiner (2001) on this point, and in the judgment that Wittgenstein was in this case led astray. Yes, there were those who took Gödel's proof as justifying platonism, but that was never a valid inference and it was not seen as such by most of the logicians who actually verified and extended Gödel's reasoning (Kleene, Turing, Rosser, Post, *et al*)—who had all published relevant work by the time Wittgenstein was writing on this topic in 1937.[26]

Wittgenstein doesn't mention platonism in his remarks on Gödel, by name or by description, but the source of his worry seems clear: the idea that if the truth of Gödel's "true but unprovable" sentence were not relativized (and disarmed) as assertibility in a "different game" or "another system", it would have to be a platonistic truth, which could not be allowed.

On this matter of the truth of Gödel's sentence I'd like to recruit some help from Torkel Franzén, who articulates what is for my money a very sane view.

> Very often in discussions of the incompleteness theorem it is regarded as unclear what might be meant by saying that an arithmetical statement which is undecidable, say in PA, is true. What, for example, are we to make of the reflection that the twin prime conjecture may be true, but undecidable in PA? In saying that the twin prime conjecture may be true, do we mean that it may be provable in some other theory, and if so which one? Do we mean that we may be able to "perceive" the truth of [this] conjecture, without a formal proof? Or are we invoking some metaphysical concept of truth, say in the sense of correspondence with a mathematical reality? (Franzén, 2005, pp. 28–29)

While recognizing that this question is "a natural one", Franzén's answer is "none of the above". Mathematicians, he says, "easily speak of truth", as in "If the generalized Riemann hypothesis is true. . . " or "There are strong grounds for believing that Goldbach's conjecture is true". In this context, he continues,

> the assumption that an arithmetical statement is true is not an assumption about what can be proved in any formal system, or what can be "seen to be true," and nor is it an assumption presupposing any dubious metaphysics. Rather, the assumption that Goldbach's conjecture is true is exactly equivalent to the assumption that every even number greater than 2 is the sum of two primes. Similarly, the assumption that the twin prime conjecture is true means no more and no less than that are infinitely many primes p such that $p + 2$ is also a prime, and so on. (Franzén, 2005, p. 29)

In each case, saying that such-and-such a mathematical proposition is true is simply another way of stating the content of the proposition. It is, Franzén says, "a mathematical statement, not a statement about what can be known or proved, or about any relation between language and a mathematical reality."

In the special case of the Gödel sentence, one might say that it can be "seen" to be true—or rather, seen that its truth is implied by consistency of S—in its metamathematical reading. The number-theoretic reading A is opaque to us, but we know that it is Goldbach-like and, as per Franzen's account, saying that it's true is really just a matter of saying "what A says" about the natural numbers (whatever exactly that may be).

Returning to Wittgenstein's "different game" claim, it seems to me that Feferman's "reflection principle" (1998, pp. 17–18) can ground the idea floated above, namely that Gödel's arithmetization of syntax amounted to a legitimate *extension* of "Russell's game"—or more generally, of a formal

[26] Alonzo Church is sometimes described as a platonist but he was surely not a platonist of Gödel's stripe, nor was he influenced in a platonistic direction by Gödel's work in particular.

system employing the Peano axioms. If we accept a system S as a representation of elementary arithmetic we can ask what commitments we thereby take on board, directly or indirectly. Clearly we're committed to the truth (or assertibility, if you prefer) of the axioms of S as well as its theorems. But if someone comes along and constructs a Gödel sentence for S, are we not committed to its truth (assertibility) too? If it were false that would imply that S is inconsistent, so we should have to revise all our commitments in respect of S.

8.4 Defences and counters

Having delivered a negative verdict on Wittgenstein's remarks on Gödel, it behoves me to say something in response to defences of Wittgenstein.[27] Much of the writing in this vein strikes me as lacking in substance, offering paraphrase of the master and assertion rather than clear arguments against the points made by Wittgenstein's critics. I will confine myself to what I find to be the most substantive defence, that of Juliet Floyd and Hilary Putnam (2000).

Floyd and Putnam (hereinafter FP) focus on §8 in Appendix I of RFM, Wittgenstein's "notorious paragraph" per Floyd (2001). To set the scene, recall that Wittgenstein had ascribed to Gödel the following argument: "Must I not say that [the Gödel sentence] on the one hand is true, and on the other is unprovable? For suppose it were false; then it is true that it is provable. And that surely cannot be!" FP zoom in on Wittgenstein's follow-up:

> 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.

FP refer to the Gödel sentence as P, following Wittgenstein, and in this context I'll do the same. I argued above that Gödel did *not* mean by "suppose it is false" that the negation of P was provable in the system in question. However, one could argue that in assessing §8 it's appropriate to follow Wittgenstein's logic—so, what *would* follow if $\neg P$ could be proved in S?

Wittgenstein states that if $\neg P$ could be proved, the interpretation of P as meaning "P is not provable" would have to be given up. FP say that this statement contains "a philosophical claim of great interest", one that has been missed by other commentators and which casts doubt on the received view that he misunderstood Gödel. Their argument gets quite technical in places, but I'll try to convey the gist.

For Gödel, the context in which the possibility of proving $\neg P$ arises is his undecidability claim: to clinch this he needs to show that neither P nor $\neg P$ can be proved. He could rule out the provability of P by assuming simple consistency of S, but to rule out the provability of $\neg P$ he needed the stronger assumption that S is ω-consistent (see Section 5.2 above). If the stronger assumption is granted it follows that $\neg P$ cannot be proved; conversely, provability of $\neg P$ would imply that S is ω-inconsistent. Grant Gödel his stronger assumption and Wittgenstein's point is moot, but FP in effect explore the implications of denying that assumption.

> To see that Wittgenstein is on to something here, let us imagine that a proof of $\neg P$ has actually been discovered. Assume, for the time being, that [S] has not actually turned

[27] To be clear, I'm just talking about defences of his critique of Gödel; I don't see him as in need of defence in most other respects.

> out to be inconsistent, however. Then, by the first Incompleteness Theorem, we know that [S] is ω-inconsistent. But what does ω-inconsistency show? ω-inconsistency shows that a system *has no model in which the predicate we have been interpreting as* 'x is a natural number' *possesses an extension that is isomorphic to the natural numbers.* (Floyd and Putnam, 2000, p. 625)

It seems there's an editorial glitch here, in that "the predicate we have been interpreting as 'x is a natural number' " makes its first appearance in this very quotation; "we" haven't been interpreting any such predicate so far. However, it becomes apparent on the following page that FP are drawing on an argument from Quine (1953), where a putative specification of what counts as a natural number has indeed been supplied, and Quine diagnoses ω-inconsistency thus:

> Where ω-inconsistency coexists with simple consistency, there is nothing outlandish afoot; there is merely a predicate, misinterpreted as "is a natural number", which has proved to be true of some other things besides the natural numbers. (Quine, 1953, p. 121)

Quine coined the term 'insegregativity' for such cases: the system in question operates with a specification of what constitutes a natural number that fails to segregate genuine natural numbers from (so to speak) 'look-alike' objects. His diagnosis would seem to make intuitive sense. If we find that a formal system proves $\forall x\, \varphi(x)$ but also $\exists x\, \neg\varphi(x)$ a possible reconciliation is that '$\forall x$' refers to the natural numbers while the contrary '$\exists x$' refers to some x that is *not* a natural number, but is mistakenly classified as such by the system in question.

I don't claim the capacity to judge this argument in formal terms, but if it's granted we get the following chain: the assumption that $\neg P$ is proved implies that S is ω-inconsistent, which implies that the specification in S of what defines a natural number is mistaken, which in turn means (per Floyd and Putnam) that the English-language interpretation of P as saying "P is not provable" is undermined:

> In short, the "translation" of P as 'P is not provable in [S]' is untenable in this case—just as Wittgenstein asserted! Floyd and Putnam (2000, p. 625)

What should we make of the FP argument? In the first place one might wonder if Wittgenstein's comment really bears such a sophisticated interpretation. The simplest interpretation would go like this: since P purportedly says 'P is not provable', if $\neg P$ were proved that would mean that '$\neg$(P is unprovable)' is proved, or in other words 'P is provable' is proved. And if one wanted to maintain that P is true one would indeed have to give up the interpretation on which it says 'P is not provable'.

FP seek to fend off a deflationary reading of §8 with the aid of R. L. Goodstein, whom we have already met in Section 8.1. Despite giving Wittgenstein's remarks on Gödel a decidedly negative review in 1957, Goodstein credited him with "remarkable insight" into certain matters of mathematical logic. He said as much in his review of RFM, and returned to this point at greater length in a paper of 1972 from which FP quote.

> I do not think Wittgenstein heard of Gödel's discovery before 1935; on hearing about it his immediate reaction, with I think truly remarkable insight, was to observe that

> it showed that the formalization of arithmetic with mathematical induction and the substitution of numerals for variables fails to capture the concept of natural number, and the variables must admit values which are not natural numbers. For if, in a system $\mathscr{A}$, all the sentences $G(n)$ with n a natural number are provable, but the universal sentence $(\forall n)G(n)$ is not, then there must be an interpretation of $\mathscr{A}$ in which n takes values other than natural numbers for which $G(n)$ is not true (in fact in 1934 Th. Skolem had shown that this was the case, independently of Gödel's work). Goodstein (1972, p. 279)[28]

(As above, Goodstein is talking about an ω-inconsistent system $\mathscr{A}$.) In his 1957 review Goodstein wrote of "the mystery that what Wittgenstein said [about Gödel] in 1935 was far in advance of" what appears in RFM (Goodstein, 1957, p. 551). FP suggest that Goodstein was missing something; on a charitable reading the "notorious paragraph" might be seen as *consistent* with the sophisticated understanding Wittgenstein had apparently evinced in 1935.

Besides Goodstein's testimony, FP refer us to another of Wittgenstein's students, Alister Watson. Watson published in *Mind* a sure-footed paper on "Mathematics and its Foundations", in which he says that his treatment of Gödel's incompleteness result "owes much to lengthy discussions with a number of people, especially Mr. Turing and Dr. Wittgenstein" (Watson, 1938, p. 445). Watson's text raises the issue of ω-inconsistency and connects it with the point that "the notion of a cardinal number, *i.e.* of a number in the everyday sense, is something that cannot be completely expressed in the axiomatic system, and is essentially non-formal" (p. 447). Moreover, FP report that Watson's paper earned (a rare instance of) praise from Wittgenstein.

In fairness, then, it's possible that Wittgenstein had a Skolem-type point in mind when he wrote §8, even though there's no hint of that in the text of RFM. Goodstein, it might be noted, goes on to say that "failure to capture the concept of natural number" does not feature in Wittgenstein's discussion; rather he "concentrates upon the meaning of *true* when one says that Gödel discovered a true but unprovable sentence." Goodstein further opines that "Wittgenstein was misled by the use of the word *true* in this connexion." (Goodstein, 1972, p. 279)

Two more points are worth making in relation to Floyd and Putnam.

First, their discussion is premised on the idea that a proof of the negation of a Gödel sentence might be found, which is counter to Gödel's explicit assumption that the system he's dealing with is ω-consistent. One might wonder whether Gödel is really entitled to that assumption, but Rosser's strengthening of Gödel's proof showed that it is not strictly necessary: the assumption of simple consistency suffices, if we're willing to substitute a somewhat more complex "true but unprovable" sentence (see Section 5.2). The relevance of this point depends on how exactly we frame the topic under discussion. If we're just talking about Wittgenstein's understanding of Gödel's proof as of 1937, Rosser is not to the point.[29] If we're asking more broadly whether Wittgenstein's remarks are helpful in thinking about Gödel's results today, Rosser's proof would seem to make Wittgenstein's remarks somewhat less interesting (no matter how charitably we read them).

[28] Goodstein's reference here is to Skolem (1933, 1934). Drawing on Vaught (1986), Skolem's 1933 result can be put this way. Let $N = \{0, 1, 2, 3, \ldots\}$ and let $\mathfrak{N}$ be a structure comprising N along with the addition and multiplication operators and the less-than relation. Then there exists a structure $\mathfrak{N}'$ *not isomorphic to* $\mathfrak{N}$ which has the same true (first-order) sentences as $\mathfrak{N}$. The non-isomorphism means that in addition to the 'shared true' first-order sentences there will be sentences true in the nonstandard model $\mathfrak{N}'$ that are not true in $\mathfrak{N}$. This issue is inherent to axiomatizations of arithmetic; it may or may not 'surface' in the form of ω-inconsistency.

[29] In principle Wittgenstein *could* have read Rosser by 1937, but it seems clear he hadn't.

Second, Timothy Bays (2004) makes a good point (actually several good points but I'll concentrate on the one I think is most telling). FP assume that if a proof of $\neg P$ were found in S, revealing that S is ω-inconsistent, there would be no choice but to give up the idea that S has the system of natural numbers (the 'standard model') as a model, and hence also give up the translation of P as "P is unprovable"—the sequel envisaged by Wittgenstein. But Bays asks, would anyone with an investment in S actually take that course? If S is *supposed to* have the natural numbers as model, would it not be much more natural to seek a fix for S to achieve that object? He goes on to suggest that a fix might be sought in the axiomatization of S, perhaps in the induction schema (which I identified in Section 5.1 as the only one of the Peano axioms which seemed open to reasonable doubt). As Bays puts it,

> My guess is that mathematicians would initially focus on the uses of induction in the proof. The hope would be that some well-motivated restriction of the induction scheme would enable us both to restore ω-consistency and to understand *why* our initial scheme went wrong (for example, perhaps we allowed induction on some subtly-paradoxical predicate/formula). (Bays, 2004, p. 204)

From this perspective, accepting that S is satisfiable only by a nonstandard model would be the last resort, not the default reaction.

At one point in RFM, in material written some years after the remarks discussed above, Wittgenstein says, "My task is, not to talk about (e.g.) Gödel's proof, but to by-pass it." (Wittgenstein, 1967, V §16) Many commentators, including FP, have quoted this sentence as if it somehow clarifies his earlier remarks, but it strikes me that in 1937 Wittgenstein signally failed to "by-pass" Gödel. His remarks from that time were clearly not intended as a formal refutation of Gödel, but all the same Wittgenstein seems determined to find *something* wrong with his proof; the question is, what exactly was wrong? As mentioned above, I think Steiner (2001) has it right: what Wittgenstein really objected to was the whiff of platonism he scented in Gödel's distinction between mathematical truth and provability. His remarks failed to hit a definite target because this distinction does not actually depend on a platonistic premise. Or so I have argued. The points made by Floyd and Putnam are surely of interest but they don't lead me to materially revise my take on Wittgenstein's remarks on Gödel.

References

Bays, T. (2004) 'On Floyd and Putnam on Wittgenstein on Gödel', *The Journal of Philosophy* 101(4): 197–210. URL https://www.jstor.org/stable/3655690.

Bernays, P. (1959) 'Betrachtungen zu Ludwig Wittgensteins "Bemerkungen über die Grundlagen der Mathematik"', *Ratio* 2(1): 1–22. URL https://www.phil.cmu.edu/projects/bernays/Pdf/bernays23_2003-05-19.pdf. Translated as 'Comments on Ludwig Wittgenstein's Remarks on the Foundations of Mathematics'.

Berto, F. (1996) *There's Something About Gödel: The Complete Guide to the Incompleteness Theorem*, Chichester: Wiley-Blackwell.

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer.

Davis, M. (1973) 'Hilbert's Tenth Problem is unsolvable', *American Mathematical Monthly* 80(3): 223–269. URL https://doi.org/10.1080/00029890.1973.11993265.

Dennett, D. C. (1995) *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, New York: Simon & Schuster.

Diamond, C. (ed.) (1976) *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge 1939*, Ithaca, New York: Cornell University Press.

Dummett, M. (1959) 'Wittgenstein's Philosophy of Mathematics', *Philosophical Review* 68(3): 324–348. URL https://doi.org/10.2307/2182566.

Feferman, S. (1998) *In the Light of Logic*, New York: Oxford University Press.

——— (2006) 'Provenly unprovable', *London Review of Books* 28(3). URL https://www.lrb.co.uk/the-paper/v28/n03/solomon-feferman/provenly-unprovable.

Floyd, J. (1995) 'On saying what you really want to say: Wittgenstein, Gödel and the Trisection of the Angle'. In J. Hintikka (ed.), *From Dedekind to Gödel*, pp. 373–426. Kluwer Academic.

——— (2001) 'Prose versus proof: Wittgenstein on Gödel, Tarski and truth', *Philosophia Mathematica* 9(3): 280–307. URL https://doi.org/10.1093/philmat/9.3.280.

Floyd, J. and H. Putnam (2000) 'A note on Wittgenstein's "notorious paragraph" about the Gödel Theorem', *The Journal of Philosophy* 97(11): 624–632. URL https://www.jstor.org/stable/2678455.

Franzén, T. (2005) *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*, Wellesley, MA: A. K. Peters.

Gerrard, S. (1996) 'A philosophy of mathematics between two camps'. In H. Sluga and D. G. Stern (eds.), *The Cambridge Companion to Wittgenstein*, pp. 171–197. Cambridge University Press.

Gillies, D. A. (1982) *Frege, Dedekind and Peano on the Foundations of Arithmetic*, Assen, The Netherlands: Van Gorcum.

Gödel, K. (1930) 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349–360. Translated as 'The completeness of the axioms of the functional calculus of logic', in Gödel (1986a).

——— (1931) 'Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173–198. Translated as 'On formally undecidable propositions of *Principia mathematica* and related systems I', in Gödel (1986a).

——— (1944) 'Russell's mathematical logic'. In P. A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, pp. 119–141. Evanston, Illinois: Northwestern University Press.

——— (1986a) *Kurt Gödel, Collected Works, Volume I*, Oxford: Oxford University Press. Edited by Solomon Feferman.

——— (1986b) 'On undecidable propositions of formal mathematical systems'. In S. Feferman (ed.), *Kurt Gödel, Collected Works, Volume I*, pp. 346–371. Oxford: Oxford University Press. Lecture notes taken in 1934 by Stephen C. Kleene and J. Barkley Rosser.

——— (1995) ‘The present situation in the foundations of mathematics’. In S. Feferman (ed.), *Kurt Gödel, Collected Works, Volume III*, pp. 45–53. Oxford: Oxford University Press.

Goldstein, R. (2005) *Incompleteness: The Proof and Paradox of Kurt Gödel*, New York: W. W. Norton.

Goodstein, R. L. (1957) ‘Review of wittgenstein’s Remarks on the Foundations of Mathematics’, *Mind* 66(264): 549–553. URL https://www.jstor.org/stable/2251066.

——— (1972) ‘Wittgenstein’s philosophy of mathematics’. In A. Ambrose and M. Lazerowitz (eds.), *Ludwig Wittgenstein: Philosophy and Language*, pp. 271–286. London: Allen and Unwin.

Hacking, I. (2014) *Why Is There Philosophy of Mathematics At All?*, Cambridge University Press.

Hilbert, D. (1926) ‘Über das Unendliche’, *Mathematische Annalen* 95: 161–190.

Hilbert, D. and P. Bernays (1939) *Grundlagen der Mathematik*, Berlin and New York: Springer-Verlag.

Keynes, J. M. (1933) *Essays in Biography*, London: Macmillan.

Kienzler, W. and S. Sunday Grève (2016) ‘Wittgenstein on Gödelian ‘incompleteness’, proofs and mathematical practice: Reading *Remarks on the Foundations of Mathematics*, Part I, Appendix III, carefully’. In S. Sunday Grève and J. Mácha (eds.), *Wittgenstein and the Creativity of Language*, pp. 76–116. London: Palgrave Macmillan.

Kleene, S. C. (1936) ‘General recursive functions of natural numbers’, *Mathematische Annalen* 112(5): 727–742. URL https://doi.org/10.1007/BF01565439.

——— (1952) *Introduction to Metamathematics*, Princeton, NJ: Van Nostrand.

——— (1986) ‘Introductory note to *1930b*, *1931* and *1932b*’. In S. Feferman (ed.), *Kurt Gödel, Collected Works, Volume I*, pp. 126–141. Oxford: Oxford University Press.

Kreisel, G. (1958) ‘Review of wittgenstein’s Remarks on the Foundations of Mathematics’, *The British Journal for the Philosophy of Science* 9(34): 135–158. URL https://doi.org/10.1093/bjps/IX.34.135.

——— (1998) ‘Second thoughts around some of Gödel’s writings: “a non-academic option”’, *Synthese* 114(1): 99–160. URL https://www.jstor.org/stable/20118014.

Lucas, J. R. (1961) ‘Minds, machines and Gödel’, *Philosophy* 36(137): 112–127. URL https://doi.org/10.1017/S0031819100057983.

Matiyasevich, Y. V. (1970) ‘Enumerable sets are Diophantine’, *Soviet Mathematics* 11(2): 354–357.

Mendelson, E. (1964) *Introduction to Mathematical Logic*, Princeton, NJ: Van Nostrand.

Monk, R. (1990) *Ludwig Wittgenstein, The Duty of Genius*, Harmondsworth, Middlesex: Penguin Books.

Nagel, E. and J. R. Newman (2001) *Gödel’s Proof*, New York: New York University Press. Revised edition, edited by Douglas R. Hofstadter.

Presburger, M. (1929) 'Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt'. In *Comptes Rendus du I congrès de Mathématiciens des Pays Slaves, Warszawa*, pp. 92–101. Translated title, 'On the completeness of a certain system of arithmetic of whole numbers in which addition occurs as the only operation'.

Quine, W. V. O. (1953) 'On ω-inconsistency and a so-called Axiom of Infinity', *Journal of Symbolic Logic* 18(2): 119–124. URL https://www.jstor.org/stable/2268944.

——— (1976) *The Ways of Paradox and Other Essays, Revised Edition*, Cambridge, MA: Harvard University Press.

Rodych, V. (2002) 'Wittgenstein on Gödel: The newly published remarks', *Erkenntnis* 56(3): 379–397. URL https://www.jstor.org/stable/20013132.

Rosser, J. B. (1936) 'Extensions of some theorems of Gödel and Church', *Journal of Symbolic Logic* 1(3): 87–91. URL https://doi.org/10.2307/2269028.

——— (1939) 'An informal exposition of proofs of Gödel's Theorem and Church's Theorem', *Journal of Symbolic Logic* 4(2): 53–60. URL https://doi.org/10.2307/2269059.

Russell, B. and A. N. Whitehead (1910) *Principia Mathematica*, Cambridge University Press.

Shanker, S. G. (1987) *Wittgenstein and the Turning Point in the Philosophy of Mathematics*, Albany, NY: SUNY Press.

——— (1988) 'Wittgenstein's remarks on the significance of Gödel's theorem'. In S. G. Shanker (ed.), *Gödel's Theorem in Focus*, pp. 155–256. London: Croom Helm.

Singh, S. (1997) *Fermat's Enigma*, New York: Walker and Company.

Skolem, T. (1931) 'Über einige Satzfunktionen in der Arithmetik', *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Mat.-naturv. kl.* (7): 1–28. Reprinted in Skolem (1970). Translated title, 'On some propositional functions in arithmetic'.

——— (1933) 'Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlreihe mittels eines endlichen Axiomsystems', *Norsk matematisk forenings skrifter* 2(10): 73–82. Reprinted in Skolem (1970). Translated title, 'On the impossibility of a complete characterization of the number sequence by means of a finite axiom system'.

——— (1934) 'Über die Nichtcharakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen', *Fundamenta mathematicae* 23: 150–161. Reprinted in Skolem (1970). Translated title, 'On the non-characterizability of the number sequence by means of finitely or denumerably many propositions containing variables only for numbers'.

——— (1970) *Selected Works in Logic*, Oslo: Universitetsforlaget. Edited by Jens Erik Fenstad.

Smith, P. (2007) *An Introduction to Gödel's Theorems*, Cambridge University Press.

Steiner, M. (2001) 'Wittgenstein as his own worst enemy: The case of Gödel's theorem', *Philosophia Mathematica* 9(3): 257–279. URL https://doi.org/10.1093/philmat/9.3.257.

Tarski, A. (1951) *A Decision Method for Elementary Algebra and Geometry*, Oakland, CA: University of California Press. URL https://doi.org/10.2307/jj.8501420. Prepared for publication with the assistance of J. C. C. McKinsey.

Turing, A. M. (1936) 'On computable numbers, with an application to the Entscheidungsproblem', *Proceedings of the London Mathematical Society* 42: 230–65.

——— (1939) 'Systems of logic based on ordinals', *Proceedings of the London Mathematical Society* 45: 161–228.

Vaught, R. (1986) 'Introductory note to *1934c* and *1935*'. In S. Feferman (ed.), *Kurt Gödel, Collected Works, Volume I*, pp. 376–379. Oxford: Oxford University Press.

Wang, H. (1996) *A Logical Journey: From Gödel to Philosophy*, Cambridge, MA: MIT Press.

Watson, A. G. D. (1938) 'Mathematics and its foundations', *Mind* 47(188): 440–451. URL https://www.jstor.org/stable/2250384.

Wiles, A. (1995) 'Modular elliptic curves and Fermat's Last Theorem', *Annals of Mathematics* 141(3): 443–551. URL https://doi.org/10.2307/2118559.

Wittgenstein, L. (1967) *Remarks on the Foundations of Mathematics*, Oxford: Blackwell.