published in
WITTGENSTEIN AND THE CREATIVITY OF LANGUAGE
ed. Sebastian Sunday Grève and Jakub Mácha, Palgrave Macmillan, 2016
https://doi.org/10.1057/9781137472540_4

4

Wittgenstein on Gödelian 'Incompleteness', Proofs and Mathematical Practice: Reading *Remarks on the Foundations of Mathematics*, Part I, Appendix III, Carefully

Wolfgang Kienzler and Sebastian Sunday Grève

I. Introduction

We argue that Wittgenstein's philosophical perspective on Gödel's most famous theorem is even more radical than has commonly been assumed. Wittgenstein shows in detail that there is no way that the Gödelian construct of a string of signs could be assigned a useful function within (ordinary) mathematics. — The focus is on Appendix III to Part I of *Remarks on the Foundations of Mathematics*. The present reading highlights the exceptional importance of this particular set of remarks and, more specifically, emphasises its refined composition and rigorous internal structure.

Ever since their first publication in 1956, Wittgenstein's remarks on Gödel and Gödel's proofs in the *Remarks on the Foundations of Mathematics* – especially those in Appendix III of Part I – have caused considerable controversy. However, only after 1985 did scholars first begin to develop serious attempts at making these remarks properly intelligible and also to positively appreciate their significance. In this connection, we can distinguish three main kinds of approaches: first, to follow Wittgenstein in questioning Gödel's interpretation of the Gödel sentence (as advocated by, e.g. Stuart Shanker (1988)); second, to try to make Wittgenstein's account in some way compatible with Gödel's own (as advocated by, e.g. Juliet Floyd (1995)); and, third, the approach which has Wittgenstein criticise and attack Gödel's proof itself, rather than 'merely' what these commentators consider to be its *interpretation* (as advocated by, e.g. Victor Rodych (1999)).[1]

The present chapter suggests a novel fourth way of approaching the relevant remarks. Gödel's own comment about Wittgenstein's remarks can give a first clue in this direction:

> As far as my theorem about undecidable propositions is concerned it is indeed clear from the passage you cite, that Wittgenstein did not understand it (or pretended not to understand it). He interprets it as a kind of logical paradox, while in fact it is just the opposite, namely a mathematical theorem within an absolutely uncontroversial part of mathematics, namely finitary number theory or combinatorics. (Gödel, 2003 [1972], p.133)[2]

Unlike other commentators, Gödel suggests that Wittgenstein questioned neither the truth, nor the philosophical significance of his (that is, Gödel's) proof – but rather that he asked critical questions about the kind of 'proof', or activity, that Gödel had put forth. Wittgenstein questioned the very *sense* of what Gödel was claiming. Furthermore, Gödel suspected that Wittgenstein pretended not to understand – but as will be seen, the case is even worse: Wittgenstein came to the conclusion that his ways of trying to understand came to nothing when he took Gödel's claims seriously. The best guess left was that it really was *some kind of paradox*.

In this chapter, our main interest is in the following set of questions: What are Wittgenstein's specific intentions in Appendix III, that is, which lines of thought is he testing out and analysing, and how, in detail, does he proceed in this text? In particular: what exactly is the internal structure of this group of remarks? For it is only once these questions have been answered adequately that one can meaningfully discuss the scope and value of Wittgenstein's remarks on Gödel.[3]

---

[1] By far the majority of commentators are in favour of this third kind of approach (this includes Bays (2004) and Steiner (2001)). For the purposes of the present chapter, however, this kind of approach is of only minor relevance.

[2] This passage occurs in a letter to Karl Menger from 20 April 1972.

[3] We would like to emphasise that it is our primary intention to clarify what Wittgenstein is doing in Appendix III; and furthermore that, while we do believe that this reading offers much systematic potential, we can no more than hint at some of these points within the scope of this chapter.

One of the main goals of the present chapter, then, is to make fruitful use of the fact that Appendix III is not merely a compilation of some loosely connected remarks, rough notes jotted down in disparate places and at different times, collected together only posthumously by some literary executor. Quite the contrary is true: in Appendix III, Wittgenstein conducts a rigorous and systematic investigation, as tight-knitted as some of the most celebrated passages from *Philosophical Investigations*.[4]

II. Prose and calculus: a useful distinction and its limits[5]

It can be quite tempting to interpret Wittgenstein's remarks along the lines of his distinction between 'prose' and 'calculus'. For instance, Juliet Floyd described her interpretation in an article from 2001 thus:

> As I see it, Wittgenstein is attempting to pare the dispensable heuristics surrounding Gödel's proof – including Gödel's introduction to his proof – away from Gödel's genuinely mathematical reasoning. In so doing, he is puncturing a certain conception of Gödel's theorem, certain philosophical prose which surrounds the proof, but not Gödel's proof itself. (Floyd, 2001, p.303; see also 2012, pp.252–4)

On 31 July 1935, two years before completing the text of Appendix III, Wittgenstein wrote in a letter to Moritz Schlick:

> 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. (CC, LW-Schlick-31-7-35, our translation)[6]

This letter is remarkable for at least two reasons. Firstly, Wittgenstein closely associates prose comments on Gödel's proofs with a certain effect

[4] For a discussion of some of the complexities of Wittgenstein's writing process, see Alois Pichler's chapter in this volume.

[5] It should be noted that, where necessary, we have emended standard English translations of Wittgenstein's writings cited in this chapter, sometimes without further indication.

[6] Wittgenstein does not mention Gödel explicitly (as, notably, he does not in the text of Appendix III either). It is obvious, however, that he is referring to the ongoing discourse about Gödel's proofs at the time.

of astonishment or surprise which such comments might have on us. Secondly, Wittgenstein unequivocally states that in order to evaluate the mathematical significance of a proof, one has to undertake a thorough *mathematical* analysis of it, *viz*. the *proof*, independently of any prose comments which might happen to surround it. These thoughts stand in close relation to Wittgenstein's remarks concerning the distinction between 'prose' and 'calculus', which he develops and stresses especially between 1929 and 1931.

Around 1930, Wittgenstein regards mathematics as autonomous in the following sense: he thinks of it as a pure calculus, constituted solely by its rules. This means in particular that he regards mathematics as strictly non-representative. Mathematics does not *describe* anything.[7] On the other hand, he regards comments about mathematical proofs, made in natural language, as 'prose that accompanies the calculus' (WVC, p.129). In this sense, he compares this kind of prose to 'the coat of paint' on a machine: 'With a machine it only matters that the cogs interlock but not what colour it is painted' (WVC, p.164). There are cases where mathematics becomes 'sensational'. In Wittgenstein's view, this usually rests on some kind of natural-language (prose) commentary which obscures the mathematical core – 'calculus' is, by definition, *not* sensational.[8] He suggests that we therefore disregard the prose part and take a closer look at the calculus. When we do this, the 'sensation' will evaporate. In this spirit he investigated, e.g. the cases of set theory (esp. 'Cantor's Paradise'), the logical paradoxes (see TLP, 3.333) and other issues in the foundations of mathematics. Calculus (or, for that matter, the activity of calculation) constitutes mathematics and, as such, remains untouched by philosophical interpretation. In particular, Wittgenstein thinks, it is therefore a mistake 'to believe that something *inside* mathematics might drop away because of a critique of the foundations':

> Some mathematicians have the right instinct: once we have calculated something it cannot drop away and disappear! And in fact, what

[7] We cannot here discuss the development of Wittgenstein's philosophy of mathematics since *Tractatus Logico-Philosophicus* in any detail. It seems worth mentioning, however, that Wittgenstein's thinking about mathematics until 1931, possibly until 1935, displays considerable continuity. A careful reading of RFM, and especially the remarks on Gödel, shows a significant change at a comparatively late stage (to be explained shortly).

[8] Of course, one could also call some things that further develop the calculus 'sensational', e.g. the invention of the decimal system, of 'zero', or of the infinitesimal calculus. But these are cases where an invention has greatly expanded the possibilities of doing mathematics. (Gödel's proof is not of this kind.)

> is caused to disappear by such a critique are names and allusions…, hence what I wish to call *prose*. It is very important to distinguish strictly between the calculus and this kind of prose. Once one has become clear about this separation, all these questions, such as those about consistency, independence, etc., will drop away. (WVC, p.149)

Wittgenstein had already advocated the same fundamental thought in the *Tractatus*, namely, that mathematics does not *describe* anything. We describe things, for example, by *using* sentences (prose) of our natural languages in order to communicate *about* the world. But, according to Wittgenstein, no such distinction between the use of symbols according to the rules of the calculus and expressing facts through some kind of predication applies to mathematics. As he already wrote in the *Tractatus*, it only leads to confusion and misunderstanding if one attempts to apply this distinction between 'doing (calculating)' on the one hand and 'describing' on the other to mathematics.[9] Just like tautologies in logic, mathematical equations 'say' nothing.

This approach has far-reaching consequences. As Wittgenstein points out, many problematic questions concerning 'consistency, independence, etc.' – and, we might add, a special sort of 'unprovability' – will simply disappear once we rid ourselves of this fundamental misunderstanding. Thus, Wittgenstein can be seen to declare his general mistrust against Gödel-style prose as early as 1931.[10]

The reason Wittgenstein stops emphasising the distinction between prose and calculus is closely tied to his later conception of language.

Starting in 1933, Wittgenstein begins to develop the conception, and related methods, of language-games. In this connection, he begins to focus on the various interrelations between linguistic and non-linguistic activities. In the course of this shift of focus, the distinction between calculus, regarded as pure (non-descriptive) activity, and prose, *mere* talk as it were, loses its seemingly fundamental importance. The conception of language-games is significantly more complex than the earlier distinction between prose and calculus (which was, by comparison, too clear-cut and simple

[9] In the same spirit, Wittgenstein emphasises the equivalence of 'process and result' in TLP, 6.1261. (See also our discussion of surprises in mathematics in Section IV, below.)

[10] The conversation quoted above (WVC, p.149) took place on 1 January 1931. Gödel's original article was only published later that same month. But Gödel had already communicated his results to several members of the Vienna Circle and others in 1930 (see also Gödel (1930)).

to be accurate). Consequently, Wittgenstein's primary interest is no longer whether some particular item can be classified as either calculus or prose, that is, as either (non-linguistic) activity or (linguistic) description of that activity. On the contrary, Wittgenstein is now primarily investigating the ways in which linguistic activity and non-linguistic activity are intimately intertwined (i.e. beyond practical separation).

With regard to Gödel, the attempt to separate prose from calculus would rest on the mistaken assumption that Gödel, in fact, *first* calculated his mathematical results (as it were, blindly), and that it was only *afterwards* that he attempted to make sense of them in ordinary language. Actually, however, the essential parts of Gödel's proof *were inspired by his philosophical agenda*.[11] Thus, 'prose' and 'calculus' cannot be neatly separated, in this case, because the calculus is designed to support a philosophical idea (in prose) – and the resulting proof is in turn explained in terms of the underlying philosophical notions. In this case, therefore, the calculus does not have a mathematical environment where it could stand on its own two feet.

If the mathematical could be separated from Gödel's commentary in accordance with the distinction between prose and calculus, *then* Wittgenstein really should have examined the proof series themselves ('from A to Z') rather than paying so much attention to Gödel's comments, which, at the end of the day, could 'drop away'. However, as we know, quite the contrary is true. Wittgenstein in fact directs *all* of his attention onto Gödel's prose, and appears to take it remarkably seriously: persistently, Wittgenstein tries out possible mathematical scenarios in an attempt to spell out concrete implications of Gödel's claims. Now, the particular reason why Wittgenstein does this lies precisely in the fact that he sees Gödel's 'commentary' and Gödel's formal proofs as fundamentally intertwined.[12]

When Wittgenstein uses the word 'prose' in some later remarks on mathematics, he is not invoking his earlier distinction between prose and calculus. Rather, he is discussing a different, although not entirely unrelated, problem: namely, in his own words, 'the curse of prose, and

[11] In a letter to Hao Wang, Gödel in fact explained that only his objectivist philosophical views led him to his results from 1930/31. See Gödel, 2003 [1967], p.398. See also Feferman (1984).

[12] It would therefore be misleading to continue to speak in the following of Gödel's prose as either 'prose' in the technical sense described above or, indeed, 'commentary'. We shall henceforth use the more neutral term 'explanations' instead.

particularly of Russell's prose, in mathematics' (RFM VII, §41). He elaborates as follows:

> The curse of the invasion by mathematical logic into mathematics is that now any sentence can be represented in a mathematical notation, and this makes us feel obliged to understand it. Although of course this notation is nothing but the translation of vague ordinary prose. (RFM V, §46)

The issue here does *not* concern fuzzy prose commenting on exact proofs. Instead, Wittgenstein is discussing the translation of ('vague') ordinary prose into logico-mathematical notation which *appears* to be exact.

Thus, in general, Wittgenstein distinguishes between the following two phenomena: first, mathematical results that are difficult and unfamiliar, and which can only be appropriately evaluated once we have thoroughly studied the relevant details; and, second, mathematical proofs that are designed to create an appearance of paradox or surprise—hence, something entirely non-mathematical and comparable instead to some kind of trick, as we shall see more clearly below. Therefore, the thorough study of any particular case of this second kind of phenomenon will be conducted according to one of the following two approaches. For we can either seek to discover the source of the appearance of paradox or surprise, and thus discard the illusion, or patiently describe how it is that mathematical practice has no room for paradoxes or surprises of this kind. In Appendix III, Wittgenstein takes the latter path.

III. The text and context of Appendix III

The text of Appendix III constitutes the only sustained discussion of Gödel in the entire *Nachlass*, which Wittgenstein himself took care to revise, revisit and edit. Appendix III reproduces the text of a typescript (TS 223) dating from 1938, which Wittgenstein separated from a more extensive typescript (TS 221) from the same year. The latter contains the entire material which makes up Part I of RFM, including all three appendices.[13] Furthermore, all relevant passages in TS 223 can be traced

[13] Appendices I and II were first published in English only with the third, revised and reset edition of 1978. The English translation of the text of *Remarks on the Foundations of Mathematics* was first prepared by G. E. M. Anscombe in 1956. The 1978 edition contains the typographically identical German and

back to one manuscript passage (in MS 118).[14] Thus, in MS 118, TS 221 and TS 223, Wittgenstein carefully prepared three subsequent versions of this same material.[15]

MS 118 also contains some additional remarks which did not make it into the final version, but which belong in the direct vicinity of the train of thoughts in Appendix III. Due to their excision for the selection in TS 221 and TS 223, these remarks have not received any attention to date. As this material is of considerable value in appreciating Wittgenstein's line of investigation, we shall include these passages and briefly comment on them, in footnotes, where they occur.[16]

One important feature of the text of Appendix III consists in the fact that Gödel's name is not mentioned even once. Wittgenstein treats the issue in a purely systematic manner.[17] The entire discussion evolves from a single statement, which purports to present the result of some sensation-causing proof, *viz*. 'There exists a sentence *P*, which is true but unprovable'. Wittgenstein reduces the notation of this 'sentence' to a single letter ('*P*'), thus indicating that he is not at all concerned with any subtle details of the steps in the formal proof procedure. The central question is: how are we supposed to react to such statements about proofs and provability, and what kinds of statement might be

English text (only more space between sections is introduced), except that in §17 of Appendix III the words 'fear and awe of mathematicians' are replaced by 'dread and veneration by mathematicians'.

[14] The relevant remarks in MS 118 are dated 22–24 September 1937. This manuscript already contains the complete text of Appendix III in its final arrangement, with the sole exceptions of the final section (§20, which stems from MS 159, pp.24r–25) and the final paragraphs of sections 14 and 15 respectively.

[15] MS 118, TS 221 and TS 223 have all been published in the *Bergen Electronic Edition*. TS 221 is also available in the exemplary *Kritisch-genetische Edition* (PU, 2001), edited by Joachim Schulte (for TS 221, see pp.329–446; for the first nineteen sections of Appendix III, including various references to MS 118, see pp.424–32). The complex textual genesis of this material resulted in several errors in its publication as Appendix III (especially in §§6, 15, 17–19), as indicated in the respective commentaries below.

[16] As is well known, Wittgenstein also wrote quite a few more remarks commenting on the 'Gödel situation'. Some of these are of great interest, but as he did not revise and order any of them in any way comparable to Appendix III, we shall have to discuss them on another occasion.

[17] Wittgenstein was never interested in anybody's thought just for the sake of getting clear about what so-and-so wanted to say, but he was always focused entirely on the systematic question that he himself was interested in.

appropriate or reasonable to make when we are confronted with certain mathematical situations?

Another important feature of the text lies in the fact that Wittgenstein does not explicitly put forward any claims of his own. The text proceeds in the form of some reminders about sentences and equations, but mostly as a dialogue (similar to many well-known passages from *Philosophical Investigations*). Wittgenstein reacts to, and enquires about, the statements of an anonymous Gödelian voice who is putting forward the claim that we have to accept a particular statement as true.

Before commenting on the remarks in detail it will be useful to say a few words about their position in the context of RFM, Part I (including its three appendices). For it is Wittgenstein's philosophy of mathematics at the time of producing the typescripts underlying this first part of the edited volume which constitutes the primary context that will help to make Appendix III more intelligible. To date, such a contextualisation has not yet been undertaken.

Wittgenstein originally intended to publish RFM I as the second instalment of his *Philosophical Investigations*. More precisely, he had planned the text of RFM I to provide the direct continuation of an earlier version, which corresponds to sections 1–188 of the *Investigations*. The main text of RFM I concerns central mathematical notions such as inference, proof, calculation and logical compulsion. The three appendices go on to deal with issues which do not concern the core areas of mathematics. In particular, all three appendices discuss problems regarding the relation between mathematics and forms of representation (including, in particular, ordinary language).

In Appendix I, Wittgenstein introduces two different negation signs. He stipulates that double negation using one of the signs ('non') yields affirmation, while double negation using the other ('ne') yields amplified negation. This leads to the apparent paradox that 'non' and 'ne' have, in some sense, the same and, in some other sense, a different meaning. He points out that the appearance of paradox is a mere symptom of one's entanglement in our concepts and notations; for, e.g. '[w]hoever calls "$\sim\sim p = p$" ... a "necessary sentence of logic" (not a stipulation about the method of presentation that we adopt) also has a tendency to say that this sentence proceeds from the meaning of negation.' (Appendix I, §11) Hence, this is one example of how the striving for meaning can lure us onto the wrong tracks in philosophy, and in the philosophy of logic in particular.

Wittgenstein then asks whether these two negation signs should be said to differ in meaning, and, if so, how exactly. He turns to questions concerning the use of these signs or rules, and how it might be decided 'what is an essential and what is an inessential, accidental feature of the notation' (Appendix I, §18). In analogy with the game of chess, he points out: 'The game, I should like to say, does not just have rules; it has a point [*Witz*].' (§20) And, he adds, if we do not see the point of a certain rule in, say, a board game, we will naturally enquire 'about the origin, the purpose, of such a rule' (§23). In concluding this first appendix, Wittgenstein finally notes the following possibility: '"If I understand the character of the game right," I might say, "this is not essential to it."' (§24) — Thus, Appendix I might be argued to bear significant relation to several lines of thought in Appendix III.

In Appendix II, Wittgenstein discusses the role of surprises in mathematics. In the *Tractatus*, he had already stressed that there could be no surprises in logic or mathematics, because 'process and result are equivalent' (TLP, 6.1261; see also 6.1251). Wittgenstein now interprets the appearance of surprises in mathematics as 'only a sign that unclarity or some misunderstanding still reigns' (Appendix II, §2). He writes: '"The proof has a surprising result!"—If you are surprised, then you have not understood it yet' (*ibid*.). Or, as Frege wrote (who, in turn, ascribes the formulation to his teacher Karl Snell): 'In mathematics, everything must be as clear as 2x2=4' (Frege, 1979 [1924/25], p.280); of course, such a standard of clarity leaves no room whatsoever for surprises.[18]

Once again, Wittgenstein sees the root of the problem as being closely connected with certain forms of (linguistic) representation: 'For a form of expression makes us act thus and so.' (Appendix II, §13) He offers the following diagnosis:

> ... a conception, presently dominant, which values the surprising, the astonishing, because it showed the depths to which mathematical investigation penetrates; ... As if, by means of these considerations, as by means of a kind of higher experiment, astonishing, nay the most astonishing facts were brought to light. (RFM I, Appendix II, §1)

Wittgenstein considers this conception to be fundamentally flawed, since it distorts the nature of mathematics (as we know it from everyday practice), and because it construes mathematics according to the model

[18] For useful accounts of Wittgenstein's discussion of surprises in mathematics, see Mühlhölzer (2002) and Floyd (2012).

of the empirical sciences. Surprises simply do not belong in mathematics. They cannot constitute a proper part of mathematics. Rather, they belong to that dubious periphery of mathematics, where mathematical results, which (by themselves) are of course required to be rigorous and *un*surprising in every single step, are transformed into statements of ordinary language which only then begin to sound puzzling and astonishing. But as long as this effect of surprise continues to exist, Wittgenstein thinks, we have not yet reached an adequate understanding of the mathematical situation.

Insofar as Gödel's proof has been regarded as a *paradigm* of a 'surprising' result in mathematics ever since its first publication in 1930/31, we may now see how the text of Appendix II can be interpreted as already dealing with Gödel in some sense. Notably, in his original article, Gödel himself speaks of his '*surprising* results concerning consistency proofs for formal systems' (Gödel 1986 [1931], p.151 [176], emphasis added). Again, Wittgenstein's task would thus be to clarify Gödel's explanatory statements in such a way as to dispel the appearance of surprise completely.

In Appendix III, finally, Wittgenstein makes the relation between mathematics and ordinary language the main focus of his discussion. For instance, Wittgenstein indicates that, contrary to a common view, there in fact exists only a 'very superficial' similarity between arithmetical equations and sentences of ordinary language (see esp. Appendix III, §§1–4 and 20). But the central question of Appendix III is how we are supposed to react to statements such as the following: '"I have constructed a sentence (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'. ..."' (Appendix III, §8).

Appendix III can thus best be understood as the detailed analysis of one *example* of what was already discussed at a more general level in the preceding Appendix II.

IV. Appendix III in detail

At a first glance, sections 1–4 and 20 can easily appear to be somewhat disconnected from the bulk of Appendix III. They are, however, of special importance for any adequate understanding of the text because they characterise the perspective from which Wittgenstein approaches the issues. The introductory remarks are designed to raise an initial awareness in the reader of what Wittgenstein regards as constituting a certain problematic unclarity in our common ways of thinking and speaking about mathematics that is generally neglected.[19]

In order to further facilitate the understanding of the section-by-section commentaries below, the following general observations concerning the specific stance taken by the author of the text of Appendix III may be in order.

Wittgenstein assumes no prior knowledge of the specificities of Gödel's results whatsoever on the part of the reader. He merely requires from his readers some familiarity with the basic mathematical practice of proceeding by means of proofs, as well as the kind of axiomatic formal system described in Russell's and Whitehead's *Principia Mathematica*. He thus intends to simulate a situation in which we *hear* about Gödel's proof for the first time—*not* one in which we study Gödel's proof series or even Gödel's introductory remarks.[20]

As already mentioned above, the entire discussion of Gödel's results proceeds without any explicit reference to Gödel (or, in fact, any other historical reference), but it evolves from a single statement which purports to summarise some mathematical result, *viz*. 'There exists a sentence *P*, which is true but unprovable'. Everything else in Appendix III Wittgenstein develops by way of examining variations of this statement.[21]

In particular, Wittgenstein consequently bypasses any specifics of Gödel's technical procedure. This means, notably, that he charitably assumes the most favourable conditions of complete formal correctness. Wittgenstein's focus is instead exclusively on Gödelian *explanations*, not of the internal correctness of the proof, but of the ways we could possibly place something like *P* within the existing system of mathematics.[22]

---

[19] Wittgenstein had already mentioned this in his letter to Schlick: 'If you hear that someone has proved that there must be unprovable sentences in mathematics, ... [f]or example, you do not know what a "mathematical sentence" is according to the conception of this proof.' (CC, LW-Schlick-31-7-35, our translation)

[20] In fact, Wittgenstein had detailed knowledge of Gödel's proof (including its more 'technical' details) and discussed it with mathematicians (see Floyd and Putnam (2000)). For his philosophical investigations in Appendix III, however, Wittgenstein deliberately left this kind of knowledge aside.

[21] This crucial fact about Wittgenstein's treatment of 'Gödel' in Appendix III has been completely ignored in the literature to date.

[22] As will be seen, however, some of Wittgenstein's remarks do seem to connect closely to aspects of Gödel's proof that are usually considered to be quite 'technical'. This has been taken as evidence that Wittgenstein is working on a technical level after all (see, for instance, the discussion of ω-consistency in Floyd and

The question which Wittgenstein investigates could also be formulated thus: how are we to react to such explanations, if we strive for conceptual clarity? The difficulty, which Wittgenstein repeatedly gives expression to, could be formulated thus: how can we even so much as *understand* these statements which appear to follow from Gödel's results? And the outcome of Wittgenstein's discussion might perhaps be put as follows: if we make the elementary assumption that in mathematics we do not accept anything which has not been proved – then a claim like 'In mathematics there exist true but unprovable sentences' will simply not be understood and, in this sense, be nonsensical. (We might as well ask: 'Are there unprovable theorems?') More specifically, the crucial way in which, upon careful examination, we find ourselves unable to understand such a statement is that, if the relevant 'sentence' is (as in Gödel's case it is) like *P*, where *P* reads '*P* is unprovable', we face the following dilemma: given our elementary assumption that mathematics is a practice which consists entirely of proofs, in order for *P* to be a proper part of mathematics, we will have to actually prove it; however, once *P* has been proved, the statement that it was 'unprovable' becomes problematic.

The dialectic of Wittgenstein's dialogues in Appendix III is designed to bring out this dilemma. It is essential to note that Wittgenstein does not, however, attempt to ultimately solve it. Rather, he carefully points out that it is only the original idea of 'unprovable sentences' that forces the dilemma upon us.

Besides bypassing all technical details of Gödel's proof procedure, Wittgenstein further refrains from discussing metamathematical or semantic considerations. Many commentators have considered this as a serious omission, invalidating most of what Wittgenstein writes. But such an objection fails to acknowledge the fact that, for the sake of the argument, Wittgenstein assumes solely the most elementary point, namely that engaging in the activity of mathematics means engaging in an activity of constructing proofs, as a shared commitment at the basis of his discussion.[23] At the heart of this practice lies the distinction

Putnam (2000)). We, however, would suggest that these points of contact should be read in the opposite direction: they can be regarded as evidence that Gödel, when he worked out the technical details of his proof, was trying to accommodate some of the issues Wittgenstein discusses in Appendix III. The point of contact is at the conceptual, not the technical, level.

[23] As he had already written in his letter to Schlick: 'For, on the other hand, we can distinguish a [practice of] mathematics in which there are *no* unprovable sentences, e.g. elementary arithmetic.' (CC, LW-Schlick-31-7-35, our translation)

between that which has been proved and that which has not been proved. It seems likely, however, that Wittgenstein regarded Gödel's metamathematical and semantic considerations as entirely separate from this basic mathematical practice and thus as, in an important sense, gratuitous.[24]

As Wittgenstein had already written to Schlick (as cited in Section I above), he finds 'not yet *anything* astonishing' in Gödel's explanations of his results. This is because, on close examination of those explanations, Wittgenstein felt quite unable to make any definite *sense* of them which could have established, in a comprehensible manner, an unambiguous connection with our basic mathematical practice of proceeding by means of proofs. Thus, as has already been mentioned, Gödel was quite right to suspect that Wittgenstein also 'pretended not to understand it' (Gödel, 2003 [1972], p.133). For, as has already been pointed out above (in Section III) and as we shall see in even more detail below: Wittgenstein does not present his case apodeictically, or even in the form of a single line of argument. Rather, in Appendix III – in the context of a discussion concerning the problematic analogy between mathematical 'sentences' on the one hand and ordinary statements on the other (see esp. §§1–4 and 20) – Wittgenstein conducts a number of detailed dialogues with an anonymous Gödelian, in which he acts out, one after the other, various possible ways in which one might *try* to give a clear sense to the kind of explanations Gödel gave of his results.[25]

---

[24] Gödel expresses the heart of his metamathematical reasoning thus: 'The method of proof just explained can clearly be applied to any formal system that, first, when *interpreted* as representing a system of notions and propositions, has at its disposal *sufficient means of expression to define* the notions occurring in the argument above (*in particular, the notion "provable formula"*) and in which, second, every provable formula is *true in the interpretation considered*.' (1986 [1931], p.151 [175–6], our emphasis)

He expresses the outlines of his semantic reasoning in the remarkable footnote 48a to his original article: 'As will be shown in Part II of this paper, the true reason for the incompleteness inherent in all formal systems of mathematics is that the formation of ever higher types can be continued into the transfinite' (1986 [1931], p.181 [191]). Unfortunately, Gödel never actually wrote this 'Part II'.

[25] It should further be noted that, similar to the case of many remarks in the *Investigations*, it would often be too simple to identify text, which is not clearly identifiable with the voice of the anonymous Gödelian, with Wittgenstein's own voice.

V. Section-by-section commentary

Note on the translation

Throughout the original text of Appendix III, Wittgenstein uses the word '*Satz*' rather than, for instance, '*Proposition*' or '*Formel*' (formula). The German word '*Satz*' is ambiguous between possible meanings in English such as 'formula', 'sentence', 'proposition' and 'theorem'. In her English translation of the appendix, G. E. M. Anscombe decided to translate this one German word as 'proposition' for most of its instances and 'sentence' for a number of others. For the purposes of the present account, which stresses the unity of the composition of Appendix III as a whole, we have decided to translate every instance of '*Satz*' (or '*Sätze*') as 'sentence' (or 'sentences') instead.[26]

Our adaptation emphasises Wittgenstein's concern about the consequences of regarding the transcription of mathematics into the notation of *Principia Mathematica* (*PM*) as an 'analysis' of mathematics. Logic is essentially a game played with sentences (that bear a certain interpretive relation to propositions of natural language), but not so mathematics. The transcription of mathematics into the logical notation of *PM* constitutes a *transformation* of mathematics into a system of sentences. This fact tends to be obscured by the term 'proposition'. However, it is only such a transcription of mathematics into a 'sentence'-form which makes Gödel's talk of a 'formula', 'proposition', or 'sentence' as 'asserting its own unprovability' possible in the first place and hence so seemingly intelligible. It should, however, also be noted that Wittgenstein is not endorsing any (careless) talk of mathematical 'sentences' – he is only adopting this way of speaking in order to bring out his point of criticism.

Section 1

This initial section prepares the ground for the discussion of linguistic analogy, or assimilation, between language-games that in fact differ from each other in important respects. Wittgenstein begins by making a basic observation, namely that, unlike the case of most assertions, we would not normally speak of a question or command as being 'true or false'. But, Wittgenstein notes, it is quite easy to imagine a language in which questions and commands would standardly be 'expressed in the form of

[26] In particular, we will hence also speak of 'sentence(s)' with regard to *Principia Mathematica*.

statements'. He points to possible examples as 'forms corresponding to our: "I should like to know if…" and "My wish is that…"'. At the end of the section, Wittgenstein leaves his readers with the question what one should think of such a language in which, say, questions would always be expressed in the form of statements.

Wittgenstein's answer, which is expressed clearly by the overall dialectic of these beginning sections of the appendix, could be put as follows: the superficial assimilation (or, even, identity) of certain utterances, by way of using the same syntactical form in each case, should not mislead us into thinking that, therefore, the respective language-games really were, or had to be, of the same kind. After all, we know that questions, commands and assertions are language-games that are played quite differently from each other.

Section 2

Next, Wittgenstein comments on some related aspects of what in German are called '*Behauptungssätze*' (literally, 'assertion sentences'[27]): 'The great majority of sentences that we speak, write and read, are assertion sentences.' He points out that the act of 'assertion' ('*Behauptung*') is not usually something which we choose to *add* to certain sentences. Rather, Wittgenstein suggests, it is 'an essential feature of the game we play with' such sentences that 'the game of truth-functions is played with them'.

Wittgenstein compares this linguistic feature to one which is essential to the game of chess, namely that 'there is winning and losing in it'. He then notes that ('of course') we could invent variants of the game in which there was no winning or losing, or in which the conditions for winning and losing respectively were different.

This time, Wittgenstein does not end with an explicit question, as he does in §1. But it seems clear that, in analogy with the foregoing considerations, this latter remark strongly invites several related questions. For example, imagine that we came across people who at first appeared to be playing chess, but who as it turns out are playing something which is indeed very similar to chess as we know it but does not include any element of winning or losing. Now, would we still be inclined to think of this as a *variant* of chess, or would we perhaps rather think of it as a different kind of game altogether? Or, similarly, if the conditions for winning and losing were different from those in the game of chess as we

[27] This is our translation. Anscombe has 'statement sentences'.

know it, *how* different would these conditions have to be in order for us not to speak of 'winning' or 'losing' anymore?

Section 3

In a brief interlude, Wittgenstein gives us another instance of a possible misunderstanding of linguistic transformation or assimilation. The kind of misunderstanding he has in mind here consists in mistaking the fact that we can theoretically distinguish between a speech act (e.g. 'commanding') and its content for the implication that the speech act is really an additional act to a preceding one, namely one of 'proposing' (or 'assuming')[28] the content as that which is to be acted upon.[29]

Section 4

At this point, Wittgenstein shifts the focus of the discussion towards arithmetic. He raises the question whether arithmetic could be done without ever even noticing the similarity between certain arithmetical symbols and the sentences of our ordinary language. Thus, he stresses the ways in which mathematics and ordinary language are also different from each other – so much so, indeed, that it could seem just as natural *not* to speak of mathematical *sentences*.

Wittgenstein then considers one natural 'point of connexion'. For someone might argue that, after all, we displayed the same gestures of approval and disapproval – for instance, by nodding or shaking our head – in the case where someone shows us their calculation results just as in the case where someone tells us about the weather, which would suggest that in arithmetic we are dealing with things which can be right or wrong, hence also true or false, just like our ordinary sentences. 'But', goes the response Wittgenstein sets against this thought, 'we also make gestures to stop our dog, e.g. when he behaves as we do not wish'—and

[28] These are our translations. Anscombe's translation of '*Vorschlag*' and '*Annahme*' as 'proposal' and 'assumption' respectively fails to capture adequately the possible meaning of the German nouns as referring to the respective *acts*, rather than their results.

[29] Wittgenstein discusses a similar case in *Philosophical Investigations*, §22, where he directly criticises Frege's account of assertion, and gives the following short example: 'We might very well also write every assertion in the form of a question followed by an affirmative expression; for instance, "Is it raining? Yes!" Would this show that every assertion contained a question?' (PI, §22). This point is also closely related to matters discussed in Appendix I.

surely, we would not want to conclude that therefore our dog's behaviour had to be of some sentence-like nature.

Next, Wittgenstein considers the linguistic analogy that seems to establish the notion that mathematics does consist of sentences (i.e. entities that can be true or false).

The analogy consists in the fact that we are used to reading a mathematical equation such as '2 x 2 = 4' as '2 times 2 is 4', whereby – and specifically by the use of the word 'is' – the mathematical equation is made to *sound* just like a sentence. However, as Wittgenstein remarks, this is 'a matter only of a very superficial relationship'. And, indeed, this sort of similarity would disappear entirely if we were to perform addition and multiplication using only an abacus.[30]

[30] MS 118 contains the following two additional sections (the additional remarks are numbered so as to indicate their relative position within the sequence of remarks as published in Appendix III):

> [§4a.] Where, in Euclid, we read: this and that is to be constructed, and in the end we have 'q.e.c.', we could also put: it is to be proved that this is the construction of that figure, and in the end we could then put 'q.e.d.', i.e. we could transform the result into a sentence that has been proved. (MS 118, p.107v, our translation)

In Euclid's *Elements*, there are sentences, usually called *theorems*, that are proved, i.e. 'demonstrated' (q.e.d.), and *problems* where some construction is to be performed. When we have solved such a problem, we can write '*quod erat faciendum*' (q.e.f.). (The abbreviation 'q.e.c.' – for '*construendum*' – seems to be Wittgenstein's invention.) Wittgenstein points out that we can easily assimilate the two different kinds of task by a change in expression. The difference in what is to be *done* in each case will, however, not be affected by such a transformation.

Wittgenstein adds two more examples:

> [§4b.] Consider the use of the form of statements when the rules of a game state: 'We arrange the pieces in such and such an order.' Imagine somebody asking: 'Is this true or false?'.
>
> I hear that it is 100 kilometres from this town to that one, and I say: '100 km – that is far. –' (A sentence using mathematical concepts only.) (MS 118, pp.107–108r, our translation)

The first remark again points out that giving, or stipulating, rules is something entirely different from asserting that some state of affairs holds or does not hold. Only in the second case does the question of truth make sense – disregarding the fact that the rule might be expressed in the form of statements.

The second remark offers an attempt to actually build a sentence that uses only mathematical concepts. However, if we say, outside of any particular situation, '100 km – that is far', this makes no sense whatsoever. Attributing ordinary properties to purely mathematical structures clearly has something awkward about it.

Section 5

Wittgenstein now introduces the main subject of Appendix III by citing the following question: 'Are there true sentences in Russell's system, which cannot be proved in his system?' He immediately introduces a second, clarificatory question asking what, in general, is called 'a true sentence in Russell's system'.

It should be noted at this point how all the crucial elements of the first question have been carefully prepared by Wittgenstein in the course of the four preceding sections. Thus, he has already, for instance, indicated to the reader some possible grounds for each of the following questions with which one might respond when faced with the question whether there are *true but unprovable sentences in Russell's system*. Hence, one might ask for example: In what sense do we speak of *sentences* in Russell's system? What does it mean to say that these 'sentences' are *true or false*? And (once one has answered these questions), in what sense can there be a distinction between 'true (or false) sentences' on the one hand and what has been proved (or disproved) in Russell's system on the other? Finally, how could we explain any distinction between 'truth' and 'provability' that would make any *real* difference; that is, would it have any *direct* (i.e. uncontroversial) consequences for the performance of our ordinary mathematical practice?

In the following sections, Wittgenstein goes on to consider a number of possible attempts to give a clear sense of what *might* be the point of a question such as the one posed in §5 about '*P*'. One method of his examination consists precisely in scrutinising the sense of respective utterances and attempted explanations by means of the set of questions mentioned in the preceding paragraph. Now, while (as we shall see) none of the particular attempts that Wittgenstein goes on to consider seems to ultimately succeed in making any clear sense, the reactions to these attempts which he rehearses, on the other hand, are intended to dispel the air of depth and surprise surrounding those attempts and to bring them back down to earth. This, no doubt, can be quite disappointing. This approach should, however, enable us to see more clearly

(J. J. Cale has a line: 'A hundred miles is not too far – unless you have to walk', in 'These Blues' on his 2004 album *To Tulsa and Back*. If we think of driving the distance, it does not seem too far; if we have to walk it, it may seem very far; but if we are to say something about this outside any context, it just comes out as nonsense.)

how difficult it is to produce any clear meaning by uttering the words, 'In any *Principia*-like formal system there exists a sentence *P* which is true but unprovable'. The following section, §6, is a good illustration of this.

Section 6

This section begins: 'For what does it mean to say a sentence "*is true*"? *p is true* = *p*. (That is the answer.)'.[31] As the following sentence explains, Wittgenstein is not advocating some kind of deflationary theory of truth. He merely points out that in a certain language-game, the circumstances under which we assert a sentence might be such that it does not make a difference whether we say '*p* is true' or simply say '*p*'. We could also put it like this: if we consider a linguistic item like a sentence '*p*', then *calling it 'true'* means simply that we *assert* the sentence, i.e. that we do not just say, or enounce, the words but that we *use* the sentence to make a move in our language-game. The language-game of Russell's system is of just this kind, *viz*. that it would mean nothing *extra* to say a sentence '*is true*'—nothing else, that is, other than simply *p*.

'If, then', the passage continues, in Russell's system the game of truth-functions is not *added* as anything over and above the assertion of a sentence in the system but coincides with it, our original question could perhaps be expressed more clearly in the following way: '"Under what circumstances is a sentence asserted in Russell's game?"' This question, however, is very simple to answer: a sentence which is asserted in *Principia Mathematica* appears either at the end of one of its proofs or as one of its *primitive propositions* (Pp): 'There is no other way in this system of employing assertion sentences in Russell's symbolism.'

Now, if this were what it means to speak of true sentences in Russell's system, then the answer to the initial question of §5 would obviously have to be *No*.[32]

[31] Unlike the printed edition, neither the manuscript version nor any of the typescript versions have quotation marks around the first occurrence of '*p*' in this remark.

[32] It may be of interest to also note how, in §§2–4, Wittgenstein has prepared his readers to see yet another question as potentially related to this point here. For in those sections he tried to raise awareness that there is an essential link between acts of assertion, and thus also 'the game of truth-functions', on the one hand and what we call a 'sentence' on the other. The starting question of the following section, §7, can be seen as addressing this very link.

Section 7[33]

The initial question of §5 now gets revised in the light of the considerations in §6. In particular, the question now becomes whether it might not be possible to speak of a *true* sentence in Russell's system in the sense, not of a sentence that is asserted in it, but of one which is (merely) 'written in this symbolism' while being true *independently* of Russell's system: '"But may there not be true sentences which are written in this symbolism, but are not provable in Russell's system?"' Once more, the possible reply that Wittgenstein goes on to consider is intended to earth this kind of question. Notably, this particular reply begins with the supposition that *true* in this sense might simply mean 'true in *another* system' (i.e. other than Russell's), that is, such sentences could 'rightly be asserted in another game.'[34]

However, if this is what *true* is supposed to mean here, then, again, this question would appear to have a rather straightforward answer, namely: (obviously) *Yes*. As Wittgenstein further notes, one might even be inclined to then reply: 'why should there not be such sentences?'

Wittgenstein now tries to clarify the situation by introducing an analogous case: 'Can there be true sentences in the language of Euclid, which are not provable in his system, but are true?' And, again, the answer seems obvious: 'Why, there are even sentences which are provable in Euclid's system, but are *false* in another system.'

Finally, the interlocutor points out that 'a sentence which cannot be proved in Russell's system is "true" or "false" in a different sense from a sentence of *Principia Mathematica*.' In other words, once we leave the system of *PM*, we have thereby also left behind its clear conditions for *true* and *false*.

Although Wittgenstein lets this particular dialogue end at this point, we, as readers, can easily imagine how it might continue from here. For,

33 In this section, it is particularly clear that the voice responding to the Gödelian is not necessarily to be identified with Wittgenstein's own voice. We have therefore chosen to explicitly refer to this voice, here as well as in §§8 and 10, as another (anonymous) 'interlocutor'.

34 This passage is thus of special relevance for Gödel's semantic considerations in his footnote 48a. This is particularly noteworthy since Wittgenstein introduces the standard interpretation of any such 'true but unprovable' sentence, as is strictly necessary only for Gödel's metamathematical reasoning from the alleged *actual* truth of his formula, not until the subsequent §8 of the appendix. In other words, it could be argued that in this striking feature of the textual structure of Appendix III, we find an indirect expression of Wittgenstein's thorough understanding of some of the deepest features of Gödel's original account.

surely, the Gödelian will be quite opposed to such a thought, since for Gödel the relevant sense of *true* is, of course, not one which is in any way relative to any particular system. At one point, the Gödelian, who feels misunderstood by this apparent trivialisation, exclaims that 'that's just a joke!'[35]

Section 8

The Gödelian now attempts to formulate their question in some additional detail. First, they give the following description of a possible scenario: '"I have constructed a sentence (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'. ... "'.[36]

Then, stressing the apparent compulsion to interpret *P* in this way, the Gödelian continues as follows: '" ... Must I not say that this sentence on the one hand is true, and on the other hand is unprovable? For suppose it were false; then it is true that it is provable. And that surely cannot be! And if it is proved, then it is proved that it is not provable. Thus it can only be true, but unprovable."'[37]

But the interlocutor remains unimpressed. The response in this section takes the line of thought begun in the preceding section one step further: 'Just as we ask: "'provable' in what system?", so we must also ask: "'true' in what system?"'.

35 Perhaps most importantly, this sort of semantic objectivism concerning the supposed unity of all relevant mathematical systems lies at the heart of Gödel's semantic considerations in his footnote 48a. At this point in Appendix III, however, Wittgenstein proceeds to the discussion of Gödel's considerations involving the sentence which appears to be saying of itself that it is not provable.

36 Gödel wrote analogously: 'In particular, it can be shown that the notions "formula", "proof array", and "provable formula" can be defined in the system *PM*; that is, we can, for example, find a formula *F*(*v*) of *PM* with one free variable *v* (of the type of a number sequence) such that *F*(*v*), interpreted according to the meaning of the terms of *PM*, **says**: *v* is a [un-]provable formula.' (Gödel 1986 [1931], p.147 [174], bold emphasis added)

37 Gödel wrote analogously: 'We now show that the proposition [*R*(*q*);*q*] is undecidable in *PM*. For let us suppose that the proposition [*R*(*q*);*q*] were provable; then it would also be true. But in that case, according to the definitions given above, ... (*non-prov*)[*R*(*q*);*q*] would hold, which contradicts the assumption. If, on the other hand, the negation of [*R*(*q*);*q*] were provable, then ... (*prov*) [*R*(*q*);*q*] ... would hold. But then [*R*(*q*);*q*], as well as its negation, would be provable, which again is impossible.' (Gödel 1986 [1931], p.149 [175])

Next, given that it has already been agreed between both sides that *true in Russell's system* means *proved in Russell's system*, the interlocutor continues their query concerning the sense of the Gödelian's words as follows. If (as would seem to be in accordance with the agreement just mentioned) '"false in Russell's system" means: the opposite has been proved in Russell's system', then, 'what does your "suppose it is false" mean?' And, since it would obviously follow that *supposing it is false in Russell's system* would mean just the same as *supposing the opposite has been proved in Russell's system*, we can conclude: '*if that is your assumption*, you will now presumably give up the interpretation that it is unprovable.'

Here, in order to avoid misunderstanding, Wittgenstein explicitly adds the following: 'And by "this interpretation" I understand the translation into this English sentence', i.e. the Gödelian's suggested interpretation into '"*P* is not provable in Russell's system"' as a sentence of the English language.[38]

Analogously, if the Gödelian's assumption were that *P* had been *proved* in Russell's system, then, contrary to the reasoning presented above, this would, again, not speak in favour of the intended interpretation. Rather, it would follow that 'the interpretation "*P* is not provable" again has to be given up'. For equally, the interlocutor continues, if – as the Gödelian intends to – one assumes *P* to be true, in the specified sense (*viz. proved in Russell's system*), then the suggested interpretation according to which *P* is not provable would seem to be falsified by contradiction just the same. Hence, there is no way that the Gödelian could maintain the originally suggested interpretation, unless they were to object to some of the previously agreed points.

At the end of the section, Wittgenstein finally takes up this latter possibility: following the previous line of thought, if *P* were supposed to be true (or false) in some other sense than *true (false) in Russell's system*, then this might be so quite independently of whether *P* or its

[38] Once this interpretation becomes doubtful, so too do the following fundamental parts of Gödel's explanations: his claim that 'it can be shown that the notions "formula", "proof array", and "provable formula" can be defined in the system *PM*' (1986 [1931], p.147 [174]), as well as his more general, and slightly more cautious, claim that the 'metamathematical notions (propositions)…can (at least in part) be expressed by the symbols of the system *PM* itself' (Gödel 1986 [1931], p.149 [175]) and thus, crucially, his bold, and not at all cautious, belief that 'all metamathematical arguments can just as well be carried out' (*ibid*., n.9) in Russell's system (*PM*) or in its Gödelised version in the domain of arithmetic.

opposite can be proved in Russell's system or not.[39] It is therefore not at all clear how (or indeed why) one could support the interpretation that the Gödelian wants to give to *P*. Wittgenstein then reminds us of some of the related points already discussed in earlier sections (esp. §2) by noting in parentheses: '(What is called "losing" in chess may constitute winning in another game.)'.[40]

Section 9

In an aside, Wittgenstein briefly addresses the question of the alleged identity of '*P*' and '*P* is unprovable'. He simply notes the following: 'It means that these *two* English sentences have a *single* expression in such-and-such a notation.' In other words, only the particular notation employed here makes the two sentences collapse into one.

There is much that could be said about this. Moreover, more might seem required from somebody who utters such a thing in order to express clearly what the relevant thought might be. Wittgenstein prefers to leave it at this brief kind of statement. It becomes apparent, however, that what is expressed here for consideration is the unorthodox view according to which *first* there are two distinct sentences of English, namely on the one hand '*P*' and on the other '*P* is unprovable', and *then* a notation (or formal system) is constructed in which, for some reason, both these sentences share just *one* formalised counterpart expression (rather than two, e.g. *P* and *Q*, as – all things being equal – would seem to be the normal thing to do).

---

39 Wittgenstein explicitly only discusses the possibility of *P* being 'false' in some other sense, but the implication of the converse possibility of it being *true* in some other sense is obvious.

40 MS 118 contains the following additional section:

> [§8a.] The whole question would be devoid of any interest if it did not connect to a superstition mathematicians have. And, again, it would not be worthwhile to refute this, if it were not a symptom of a widely spread 'disease of thought' [*Denkkrankheit*]. (MS 118, pp.110–111r, our translation)

Wittgenstein indicates that questions such as the one presented by the Gödelian in §8 are only of interest to his philosophical investigations insofar as they are examples of even more fundamental tendencies of thinking (such as the tendency to think that mathematics was *about* something). See his similar points in §13 and at the end of §17.

Hence, this section raises a question addressed to the Gödelian, but this line of inquiry is not carried any further at this point.[41]

Section 10

This section continues the dialogue of §§5–8. The voice of the Gödelian is now protesting against the reasoning from §8: '"But surely *P* cannot be provable, for, supposing it were proved, then the sentence that it is not provable would be proved."' Hence, the Gödelian *rejects* the suggestion that one may have to change one's initial interpretation of *P* as the English sentence '*P* is unprovable' once *P* has been proved in the relevant system in which it was constructed. Rather, *insisting* on the correctness of their initial interpretation, the Gödelian reiterates their argument from §8, *viz*.: *P* is constructed in such a way that any possible proof of it, at the same time, had to be a proof of its unprovability; since, however, (allegedly) *by means of certain definitions and transformations*, *P* can be interpreted as the English sentence '*P* is unprovable', this particular circumstance (i.e. that any proof of *P* had to be a proof of its unprovability) does not result in a mere mathematical contradiction or philosophical paradox but shows that *P* is in fact *true*, because it *says* just what it is, namely, unprovable; or, at least, this is how the original argument of the Gödelian goes.

Seeing that the Gödelian remains entirely unimpressed by the previously presented considerations in favour of changing the initial interpretation of *P* once it has been proved, the interlocutor now sketches a

---

[41] MS 118 contains the following additional section, pushing the issue of double notation still further:

> [§9a.] Now, imagine that somebody asked me: 'Is "*P*" provable?' – Now I answer: '*P*.' Obviously this is no answer; in English I would have had to answer: '"*P*" is unprovable'. But imagine somebody asked me in that other notation: '*P*?' – What am I to answer? (MS 118, p.111r, our translation)

This remark stresses the problem of jumping between the two levels of formal and ordinary expression. If somebody asks me, in English, 'Is *P* provable?', then the answer will be an English sentence, e.g. '*P* is unprovable'. We might have thought that, given the supposed equivalence of *P* and '*P* is unprovable', we could just as well have given as an answer simply: '*P*'. This, however, would either be a case of silly playing around—it would sound like it came from a Monty Python sketch—or else it would just be nonsense (i.e. it would neither constitute an affirmative nor a negative answer). However, if we take the formal notation seriously, the *question* would already fail to yield a clear sense: '*P*?' What could I answer? '*P*'? But what *could* this even mean? *P*? And would this not, again, just be silly playing around?

situation in which *P* has actually (just) been proved or, as is conceded for the sake of the argument, in which 'I believed—perhaps through an error—that I had proved it'. If, then, as a mathematician, I find myself in such a situation, the interlocutor asks, '**why** should I not let the proof stand and say I must withdraw my interpretation "*unprovable*"?'[42] Thus, Wittgenstein asks how the Gödelian would react if they were confronted with a proof of the supposedly unprovable: would they reject the proof, or would they revise their interpretation?[43]

Section 11

Wittgenstein begins to consider the possible consequences of the assumption which was first articulated in §8. He writes: 'Let us suppose I prove the unprovability (in Russell's system) of *P*; then by this proof I have proved *P*.'

In what follows Wittgenstein acts out a possible reaction to this imagined mathematical datum.[44]

First, he points out that, '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'. Thus, we simply do not know whether *P* is supposed to belong to Russell's *Principia Mathematica* or not. Wittgenstein lets a rather laconic comment follow: 'That is what comes of making up such sentences.'[45]

---

[42] All **bold** emphases in this chapter are ours.

[43] See also the related discussion in §17.

[44] It might be worth noting that the notion of truth plays no direct role in the discussion of §§10–15. In these sections, Wittgenstein limits the focus of the discussion exclusively to mathematics as a practice which consists of proofs only.

[45] It is important to note that this first possible reaction that Wittgenstein considers is not one according to which *P* was necessarily viewed as a contradiction occurring in *PM*. Rather, Wittgenstein's point is that, if there is no proof of *P* in *PM*, then it is first and foremost *questionable* whether *P* should be regarded as a (well-formed) part of *PM* or not. The question whether *P* might constitute or lead to a contradiction in *PM*, on the other hand, will only become relevant if the former question is answered in the affirmative. However, Wittgenstein makes no direct statement about how to decide the first question. Recent interpretations of Wittgenstein's stance on Gödel as being essentially the same as that of some paraconsistent logicians (see, e.g. Berto (2009)) misconstrue Wittgenstein's investigations of Gödel in Appendix III. Furthermore, the overall reading of Appendix III as presented in the present chapter would suggest that, contrary to those paraconsistent accounts, Wittgenstein was inclined to think that one should regard *P* as not being a well-formed part of *PM* rather than as constituting a contradiction in it.

This immediately prompts the objection that Wittgenstein would not be doing justice to the seriousness of the situation: 'But there is a contradiction here!' Wittgenstein agrees, but comments calmly: 'Well, then there is a contradiction here. Does it do any harm here?'

What might at first appear as a merely rhetorical question (the Gödelian would say: *yes, of course!*), actually receives some critical attention in the next section.

Section 12

Wittgenstein now explains his unorthodox view of contradictions. To him they are not dangerous but merely useless. In his view, the problem with contradictions is not that something bad would follow from them (explosion, i.e. *ex falso quodlibet*); this he calls a *superstition*.[46] Rather, a contradiction *indicates* a problem: something has gone wrong already, namely that we have got (seriously) entangled in our concepts and notations. Thus, a contradiction should make us stop and consider things anew.[47]

In obvious – although explicitly unresolved – analogy to the concluding question of the foregoing section, Wittgenstein asks whether the contradiction in '"I am lying.—So I am not lying.—So I am lying.—etc."' does any 'harm'?[48] In a second question, he then asks, more precisely, whether the usefulness of our natural language is in any way diminished by the possibility of forming such a series of contradictory sentences.

The possible answer Wittgenstein considers displays a relaxed attitude similar to the one in the preceding section, namely: 'the sentence *itself* is useless, and these inferences equally; but why should they not be made?' The only reason why we usually do not want to produce contradictions (except in jest) is their uselessness: 'It is a profitless performance!' Wittgenstein concludes the section by offering still another way of looking at it: 'It is a language-game with some similarity to the game of thumb-catching'—thus indicating that this could be a playful use of language without any serious application, but that this need not mean that it is therefore illegitimate or dangerous.[49]

---

[46] See §17.

[47] See also PI, §125.

[48] Gödel emphasises that his formula 'is closely related to the "Liar"' (1986 [1931], p.149 [175]). Furthermore, Gödel adds: 'Any epistemological antinomy could be used for a similar proof of the existence of undecidable propositions.' (*ibid.*, n.14) See also Gödel, 1986 [1934], pp.361–3 [20–2], where he demonstrates this for 'the Liar'.

[49] MS 118 contains the following parenthetical addition to §12, in which Wittgenstein briefly explains the game of thumb-catching:

Section 13

In a brief remark, Wittgenstein suggests that a contradiction such as the one discussed in the previous section is 'only' of philosophical interest in virtue of, say, its psychological and anthropological significance ('because it has tormented people'). What we can learn from this, he writes, is 'how tormenting problems can grow out of language, and what kind of things can torment us.'

The somewhat provocative implication of this remark is of course that, contrary to the beliefs of the majority of 20th century logicians, paradoxes such as 'the Liar' are of no serious logical or mathematical significance. We might as well view them as allowing for little more than logical exercises or games, rather similar in purpose to the game of thumb-catching; or thus seems to be Wittgenstein's suggestion in connection with the ending of §12.

In analogy with *P*, then, from this perspective, a formula whose proof (in Russell's system) would be equivalent to a proof of its own unprovability will seem just as *useless* as a (pseudo)sentence which appears to be asserting its own falsity (while, to be sure, such a string of words does not actually assert anything about anything – it is of no use whatsoever in the language-game of making assertions).[50]

Section 14

Wittgenstein now introduces a new line of thought, and compares the situation of *P* with unprovability proofs. The latter are not ordinary, straightforward proofs, but proofs 'concerning the geometry of proofs', as he describes it. They line up possible proofs, or constructions, and then show that a certain proof, or construction, will not appear in this line: 'Quite analogous e.g. to a proof that such-and-such a construction is impossible with ruler and compass.'

[§12a.] (This is played like this: you hold the thumb of your right hand with your left hand, so that its tip peeks out from the left hand. Now you quickly withdraw the right hand from the grip of the left and try to catch the tip of the right thumb with your right hand before it withdraws.) (MS 118, pp.112r–113, our translation)

[50] And, of course, Gödel did not intend to make any (ordinary) *mathematical* use of his formula within *Principia Mathematica*-like systems. Rather, the mere constructability – the apparent existence – of such a formula sufficed for Gödel's *philosophical* (or *metamathematical*) cause concerning what he conceived of as the supposedly transfinite totality of mathematics.

A geometrical proof of this sort, Wittgenstein notes, 'contains an element of prediction'. For example, the proof that there can be no general method for the trisection of an angle with ruler and compass can be used in order to dissuade someone from trying to find such a general method: for, in other words, it has been proved that they *will* not succeed. A geometrical proof of this sort, Wittgenstein adds, '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).'

In a separate line at the end of this section, Wittgenstein finally comments: 'A contradiction is unusable as such a prediction.'[51] This postscript sentence reconnects the side discussion of §§12 and 13 with the continuing main line of the present investigation concerning the likely details and consequences of a possible proof of *P* or a proof of the impossibility of such a proof respectively. Wittgenstein's point in this final line may be taken to be the observation that a mere contradiction such as the one mentioned in §11 could not function like the kind of *geometrical* proof of unprovability of the trisection considered in the present section, because in the former case there is no such line-up, *not* some (mathematical) *thing* which was then shown to be unprovable – but there is, rather, nothing (except the contradiction itself).[52] Thus, we see that the intended analogy never really gets off the ground (at least not as long as one insists on the Gödelian interpretation of *P* as the English sentence *P is not provable in Russell's system*).[53,54]

---

[51] This line appears as a handwritten addition in TS 223.

[52] Similarly, it is not the case that people had previously tried to prove Gödel's formula but then Gödel showed that it could not be done.

[53] Cf. the discussion in Floyd (1995) which is based on the assumption that Wittgenstein sees a strong analogy between Gödel's procedure and the proof that there can be no general method for the trisection of an angle with ruler and compass.

[54] MS 118 contains the following additional section, in which Wittgenstein attempts to elucidate the kind of difficulty addressed by his investigation—which, as he emphasises, does not so much involve a technical skill as, rather, a certain sense for philosophical difficulties:

> [§14a.] People have sometimes said to me they cannot make any judgement about this or that because they have never learnt philosophy. This is irritating nonsense, it is being assumed that philosophy is some sort of science. And people speak of it as they might speak of medicine. – What one can say, however, is that people who have never carried out an investigation of a

Section 15[55]

This section continues the reflection, begun in §14, with regard to questions concerning an unprovability proof for *P* that arise from parts of the preceding discussion starting in §8. In elaboration of the discussion in §6, Wittgenstein introduces the following point: 'Whether something is rightly called the sentence "*ξ* is unprovable" depends on how we prove this sentence.' This point is directed at the one made about *P* by the voice of the Gödelian in §8, according to which ***by means of certain definitions and transformations*** *it can be interpreted such that it says: 'P is not provable in Russell's system'*. Wittgenstein now suggests that there are reasons to believe that *definitions and transformations* alone are not sufficient to validate such an interpretation as a sentence of our natural language. On the contrary, he contends, '[t]he proof alone shows what counts as the criterion of unprovability.'

Wittgenstein then explains that *if* one wishes to speak of the '"sense"' of a sentence, one is best advised to examine 'the system of operations, of the game, in which the sentence is used'. For, he continues, the proof

> philosophical sort, like most mathematicians for instance, are not equipped with the right optical instruments for that sort of investigation or scrutiny. Almost, as someone who is not used to searching in the forest for berries will not find any because his eye has not been sharpened for such things & he does not know where you have to be particularly on the lookout for them. Similarly someone unpractised in philosophy passes by all the spots where difficulties lie hidden under the grass, while someone with practice pauses & senses that there is a difficulty here, even though he does not yet see it. – And no wonder, if one knows how long even the practised investigator, who realizes there is a difficulty, has to search in order to find it.
>
> If something is well hidden it is hard to find. (MS 118, pp.113r–114r, as translated in CV, pp.33e–34e)

[55] In the printed edition the final paragraph contains the phrase '"proof of the unprovability of *P*"'. However, it should correctly read '"proof of the unprovability" of *P*', following Wittgenstein's correction in TS 223. It should further be noted that the printed edition reproduces the typographical substitute 'X' that was used in TS 223 in place of the Greek letter '*ξ*' originally used by Wittgenstein in MS 118. (The use of the Greek letter is presumably inspired by Frege's manner of introducing notational devices. See also e.g. TLP, 5.5.)

of the sentence within this system will naturally form the centre of its systematic relations to other parts of this same system (and only the proof makes these relations fixed and precise); and, thus, if anything, Wittgenstein notes, it can be said of a *proof* of a sentence that it 'shows us its "sense".' Hence, following this thought, as Wittgenstein writes at the start of the section: without a proof, nothing 'is rightly called the sentence "ξ is unprovable"'.

We find here an important element of Wittgenstein's motivation for taking the Gödelian's argument, as presented in §8, so exceptionally seriously. There, the Gödelian argued that it could be proved, informally, that the interpretation of *P* as '*P* is unprovable' – supposedly reached *by means of certain definitions and transformations – had* to be correct because the converse assumption, i.e. *that it was proved*, leads to a contradiction.[56] However, besides other possible objections to this argument, Wittgenstein has already indicated that even a formal contradiction of *P* in Russell's system will not be sufficient to establish the required kind of unprovability proof. Returning to the discussion of possible unprovability proofs in §14, Wittgenstein concludes the present section by writing: 'Thus the question is whether the "proof of the unprovability" of *P* is here a forcible reason for the assumption that a proof of *P* will not be found.' – In §17, then, Wittgenstein provides an elaborate sketch of a possible answer to this question.

Section 16

In this section, Wittgenstein points out an important implication of the remarks in the foregoing section: namely that, in particular, even in the case of *P*, which we constructed in such a way that (supposedly) *by means of certain definitions and transformations it can be interpreted such that it says: 'P is not provable in Russell's system'*, there would still remain a question as to the exact meaning of this latter sentence once *it* has been proved.[57] Wittgenstein puts it thus: 'The sentence "*P* is unprovable" has a different sense afterwards—from before it was proved.'

[56] As has been pointed out already in the footnotes to our commentary on §8, above, Gödel used a very similar argument in his original article.

[57] E.g. this latter sentence might be proved in the form of an unprovability proof of the geometrical kind as considered in §14.

Before '*P* is unprovable' has been proved, '*what* is to count as a criterion of its truth is not yet *clear*, and—we can say—its sense is still veiled.' But it is only once this sentence has been proved – once it has been constructed as 'the terminal pattern in the proof of unprovability' – that it can have a clear sense for us. More specifically, for example, only then will we be in a position to specify just *what kind* of unprovability proof it is, the terminal pattern of which it is supposed to represent (and hence also what the resulting consequences are for the initial interpretation of *P* as the English sentence '*P* is unprovable').[58]

Section 17[59]

The section begins with the following series of questions: 'Now how am I to take *P* as having been proved? By a proof of unprovability? Or in some other way?' Wittgenstein once more revisits the Gödelian argument of §8. This present section concentrates in particular on the premise that *if P is proved, then it is proved that it is not provable*. Wittgenstein considers four possible scenarios.[60]

---

[58] Wittgenstein does not make it explicit in this section, but the context of §§14–17 clearly indicates that a proof of '*P* is unprovable', which is what is at question here, would in no way show the initial interpretation of *P* as the English sentence '*P* is unprovable' to be correct. See also our commentary on the following section.

[59] The printed edition fails to reproduce Wittgenstein's use of the symbols '~' (negation) and '⊦' (assertion), as can be seen by comparing MS 118 and TS 223. But Wittgenstein's careful employment of this notation is important because it highlights the interplay between verbal (not-*P*) and symbolic expression (~*P*).

[60] Again, it is important to note that Gödel made a very similar argument to the one presented in §8 in Gödel, 1986 [1931], pp.147–9 [174–5]; please see the quotation of the relevant passage in the second footnote to our commentary on §8, above (see also the first footnote to our commentary on §18, below). As discussed in Section IV above, while Wittgenstein's discussion is based on the elementary assumption that mathematics is a practice which consists only of proofs, Gödel was more than prepared to give up this assumption in favour of introducing the metamathematical distinction between truth and provability into mathematics. Hence, while Wittgenstein discusses the possible consequences that concrete proofs of *P* or ~*P* might or might not have for mathematical practice, Gödel in his original account proceeds solely on a general level of 'provability' with no apparent consideration of *concrete* possibilities of proof. Accordingly, the argument of §8 speaks of supposing *P* were 'proved' rather than 'provable'. Since, however, the argument of §8 proceeds entirely in a hypothetical, or modal, manner (*if it is proved…*; *Thus it can only…* etc.), the argument remains essentially the same.

1. *P has been proved by a proof of unprovability.* Wittgenstein comments on this first possible scenario only very briefly. He mainly emphasises, again, just how unclear it is what exactly such a proof might look like in this particular case. He writes: 'Now, in order to see *what* has been proved, look at the proof. Perhaps it has here been proved that such-and-such forms of proof do not lead to *P*.' This, of course, is a scenario that is simply ignored by the Gödelian, who is constructing no such systematic array of proofs at all.
2. *P has been proved in a 'direct' way.*[61] Wittgenstein notes that, of course, 'in that case there follows the sentence "*P* is unprovable"' (*viz. by means of certain definitions and transformations*). And, for reasons already indicated (esp. in §§8, 10 and 15), Wittgenstein thinks that, in such a scenario, 'it must now come out how this interpretation of the symbols of *P* collides with the fact of the proof, and why it has to be given up here.'
3. *~P is proved.* Wittgenstein immediately asks: 'Proved *how*?' And he considers the following: 'Say by *P*'s being proved directly—for from that follows that it is provable, and hence ~*P*.' This sort of reasoning makes implicit use of the initial interpretation of *P* as being equivalent to the English sentence '*P* is not provable'.[62] Accordingly, Wittgenstein suggests that, when being asked by someone '"Which is the case, *P*, or not-*P*?"', it would then be an accurate answer to say 'both' and to elaborate as follows: '"⊦*P*" stands at the end of a Russellian proof, so you write in the Russellian system: "⊦*P*"; on the other hand, however, it is then provable and this is expressed by "⊦~*P*", but this sentence does not stand at the end of a Russellian proof, and so does not belong to the Russellian system.' — In the following sentence, Wittgenstein explains further, in some detail, why it might be acceptable for us to describe the situation in this way and, in particular, why *P* and ~*P* need not be seen as constituting a contradiction in this case. Drawing

[61] Wittgenstein hastens to add, in parenthesis, 'as I should like to put it', because so far we do not know what '*P* has been proved in a direct way' amounts to; and because, as he then goes on to note, this might evoke an immediate temptation to infer ~*P* via the initial interpretation of *P* as '*P* is unprovable' (and also, perhaps, because – as we happen to know, but Wittgenstein's readers in general might not *and* need not – Gödel constructed his formula in such a way that its negation is deducible from it and vice versa).

[62] As already mentioned in the previous footnote, it would be a mistake to think that what Wittgenstein meant here was that ~*P* could be formally deduced from *P*. This is a technical detail about Gödel's actual procedure, which Wittgenstein does not discuss in Appendix III.

on the discussion in §§15 and 16, Wittgenstein points out that, it might be argued, it is only *after* the proof of *P* has been given that the sense of the initial interpretation of *P* as being equivalent to the English sentence '*P* is unprovable' becomes clearer to us; namely, once this particular proof of *P* has been given, we see that '*P* is unprovable' could at least *not* mean that '*this* proof did not exist' (hence the possible verbal answer: '*P* and not-*P*', where only *P* stands at the end of a Russellian proof, and *P* is interpreted as being *roughly* equivalent to the English sentence '*P* is unprovable', i.e. where it is not entirely clear what the *exact* meaning of this latter sentence might possibly be). — Finally, Wittgenstein adds the more general remark that such a direct proof of *P* creates a '*new situation*' which requires a *decision* from us. Either we accept the proof of *P* and, consequently, revise our initial interpretation of it (*viz*. as at least not meaning that *P* was *absolutely* unprovable); or we uphold our initial interpretation of *P*, but then we must reject the proof instead.[63]

4. *~P is proved directly*. Wittgenstein first considers the following possible reaction to this imagined mathematical datum: 'it is therefore proved that *P* can be directly proved!' This reasoning is analogous to the one in the previous scenario. In this case, it is apparently inferred from *~P*, via the initial interpretation of *P* as '*P* is unprovable', that *P* must be provable after all. Accordingly, Wittgenstein simply refers us to his previous remarks ('this is once more a question of interpretation'). — In a characteristic turn, finally, Wittgenstein suggests taking the talk of *P*'s alleged (direct) *provability* seriously and asks how one might actually react if 'we now also have a direct proof of *P*'. In a similar fashion to §§10–12, Wittgenstein simply writes: 'If it were like that, well, then it were like that.'[64] Then he concludes the present section with the following aside: '(The superstitious dread and veneration by mathematicians in face of contradiction.)'[65]

63 Incidentally, Anscombe wrongly translates the italicised '*noch*' (in '*noch einen Beweis*') as 'further' ('a *further* proof'). But the question discussed here is whether we will go on and still call this a proof or whether we will feel forced to give up this view. Therefore, this would be a more accurate translation: 'now we have to decide whether we will (*still*) call *this* a proof, or whether we will still call *this* the statement of unprovability'.

64 In this alternative translation, we have attempted to mirror Wittgenstein's tautological phrasing in the original of this passage (*viz*. '*Wäre es nun so, nun, so wäre es so.*'), which it is difficult to reproduce in grammatical English.

65 The manuscript version contains the insertion '*Sehr komisch ist*' at the beginning of the sentence. Hence, the translation might read: '(The superstitious

Section 18

Following the re-examination of one of the two main premises of the Gödelian argument from §8 in the preceding section, §17, Wittgenstein now also briefly revisits the other main premise of that argument. At the beginning of the present section, then, he has a voice reciting the central idea of the premise: '"But suppose, now, that the sentence were *false*—and hence provable?"'

As was the case with the supposition of a possible proof of *P* (the first main premise of the argument from §8), so, equally, in this case Wittgenstein takes the supposition of the falsity of *P seriously*, and asks what concrete mathematical situation might actually be described in such a way: 'Why do you call it "false"? Because you see a proof?—Or

---

dread and veneration by mathematicians in face of contradiction is very odd [or: funny].)' (MS 118, pp.116–116r)

Wittgenstein was no doubt aware that this particular remark, which in the manuscript version still constitutes a separate section, would present an exceptional provocation, but he kept it in the text (albeit in parentheses). As it turned out, Gödel himself did take offence at this sentence and wrote in his letter to Menger: 'Incidentally, the whole passage you cite seems nonsense to me. See, e.g. the "superstitious fear of mathematicians of a contradiction".' (Gödel, 2003 [1972], p.133) But note that Gödel left out the 'awe' (or 'veneration') which is also mentioned by Wittgenstein. (The 1956 translation, from which Gödel is quoting, has 'fear and awe of mathematicians'.)

Perhaps Gödel is also displaying a vague sense of the actual relevance that Wittgenstein's provocative aphorism has for his own work. For the final scenario which Wittgenstein considers in §17, before concluding with this aphorism, can be shown to be of considerable relevance for Gödel's explanations of his results in the following two ways. First, Wittgenstein's discussion indirectly suggests that Gödel made one decision rather than another, namely when Gödel concluded that his formula was not provable in the formal system in which he constructed it *rather than* that it had a proof and that its negation had one too (a similar point is also made by some proponents of paraconsistent logics). Second, Wittgenstein's discussion indirectly suggests that, if Gödel could not rely on his metamathematical reasoning, his semantic considerations would seem to have involved decision-making of a similarly *ad hoc* kind. For, if Gödel's metamathematical reasoning, according to which his formula is actually true, is rejected, then: how do we know in what direction to continue 'the formation of ever higher types…into the transfinite' (Gödel, 1986 [1931], p.181 [191], n.48a) which Gödel suggests; for example, why should we not develop it in both directions at the same time, i.e. in the direction of systems in which Gödel's formula or its negation can respectively be proved, or both? (With regard to this latter question, it is interesting to note that, in the section immediately following this one, Wittgenstein actually considers the possibility that *P* could be *false* in some system just like there can be formal systems in which not-not-*p* is not equivalent to *p*.)

for other reasons?' Assuming the latter (and having already dealt with the former possibility in §17), Wittgenstein writes that 'in that case it doesn't matter'. For, as he goes on to point out, analogously – that is, not with reference to a proof in Russell's system, but 'for other reasons'[66] – it is perfectly fine to call, e.g. the sentence expressing the law of contradiction (*PM* 3.24) or the one expressing the principle of double negation (*PM* 4.13)[67] 'false', if what we mean by that is that, say, in ordinary discourse – as it were, in *violation* of those rules of Russell's system – 'we very often make good sense' by saying 'Yes and no' or by using double negation to emphasise negation.

Thus, Wittgenstein points out that, if we make a serious attempt at understanding what exactly might be meant by expressions such as *suppose ξ is provable* or *suppose ξ is false*—or, more specifically, if we try to think of possible *concrete* mathematical situations and implications—we will find ourselves once again unable to do so.

Section 19

Wittgenstein finally turns his attention to the alleged conclusion of the Gödelian argument from §8. He writes: 'You say: "…, so *P* is true and unprovable".' Given the preceding reflections in the appendix, Wittgenstein begins his reply as follows: 'That presumably means: "Therefore ⊢*P*." That is all right with me—'. For the sake of the argument, Wittgenstein lets the Gödelian have their initial interpretation of *P* as the English sentence '*P* is unprovable' (not, however, its alleged *truth*). Instead, Wittgenstein now asks: 'but for what purpose do you write down this "assertion"?'[68]—thus implying that, as has been indicated by the beginning sections of Appendix III, that it is rather questionable what kind of language-game we may play with such a sentence at all.

In parentheses, Wittgenstein notes that its 'purpose' might be compared to that of someone's 'assertion' that it followed – more or less naturally – from certain principles of architecture that some extremely fancy chalet had to be built somewhere in an uninhabitable part of the

[66] Or, as Wittgenstein puts it in §8: 'if the sentence is supposed to be false in some other than the Russell sense'.

[67] The formula in the text of §18 should read '$p \equiv {\sim}{\sim}p$' (rather than '${\sim}{\sim}p = p$', as in the printed edition).

[68] Crucially, the printed edition erroneously leaves out the preceding *assertion* sign in '⊢*P*'.

universe.[69] – However, even if such a chalet could in principle be built there, actually building it would nevertheless remain an essentially useless exercise and waste of resources. This kind of 'assertion' would hence hardly make any sense to us. And, therefore, it would make equally little, or even less, sense to us that such an 'assertion' should be called *true*.

This last point is expressed by Wittgenstein in the form of a question in the *irrealis* mood: 'And how could you make the truth of the assertion plausible to me, since you can make no use of it except to perform those little tricks?'[70] This question concludes the detailed discussion of the Gödelian argument.[71]

Section 20

In a final remark, Wittgenstein returns to the beginning of his reflections in §§1–4.

He first reminds us 'that the sentences of logic are so constructed as to have *no* application as *information* in practice.' Therefore, he continues, one might well argue that it would be better not to call them *sentences* at all.

Then, with special regard to the construction of *P* (and, in particular, the construction of its alleged interpretation *by means of certain definitions and transformations*), Wittgenstein further points out that, 'if we append to these "sentences" a further sentence-like structure of another kind, then we are all the more in the dark about what kind of application (what kind of sense) this system of sign-combinations is supposed to have'.[72]

In other words, not only is it unclear whether we should really speak of mathematical *sentences* as such (as was pointed out in §§1–4), but even if a formal system such as that of *Principia Mathematica* presented the correct logical analysis of mathematics – as is assumed by the

[69] Wittgenstein's example of 'Mount Everest', it should be noted, was an uninhabitable part of the universe of just this kind at the time of writing (*viz*. in 1938, i.e. 15 years before it was first successfully ascended by a human being).

[70] Anscombe's translation of '*Kunststückchen*' as 'bits of legerdemain' is potentially misleading in that it suggests a rather more specific kind of action than the original German does.

[71] See also Appendix I, §§18ff.

[72] We have added in brackets a part of the sentence which Anscombe's translation (strangely) omits, *viz*. 'what kind of sense' in translation of '*was…für einen Sinn*'.

Gödelian – this would still leave open the question of whether it is reasonable to speak of *logical sentences* in the first place. The specific talk of *mathematical sentences* that the Gödelian engages in, which appears so natural, is in fact problematic in a twofold respect.

Thus, at the end of a quite thorough investigation, it finally remains very much doubtful how we could possibly interpret *P* as a sentence of the English language. But 'the mere *ring of a sentence*', Wittgenstein notes – almost trivially – in concluding the appendix, 'is not enough to give these connexions of signs any meaning.'

—As already remarked upon, it is the joint function of §§1–4 and 20 to make explicit the perspective within Appendix III from which Wittgenstein approaches the question: *Are there true sentences in Russell's system which cannot be proved in his system?* But it would be wrong to think of these five sections as something like a mere *frame* which is somehow disconnected from the central discussion of Appendix III (as, unfortunately, the standard translation can easily make it appear). On the contrary, the function of §20 could be accurately characterised as the final element of the picture, which has been carefully crafted so as to connect the picture to 'the frame', in order that, if this were not clear from the start, the frame can finally be seen to be a proper part, as it is intended to be, of the picture *itself*. It is thus also indirectly suggested – namely in virtue of the structural position of §20 within the overall textual composition of the appendix – that the reader return to the beginning of the text and read the sections once again, carefully, from the start.—For, even if not much else, we hope to have demonstrated at least this much through this commentary on the text of Appendix III: that it can hardly be read carefully enough.

VI. Concluding remarks

We have argued in this chapter that, when Appendix III is read carefully, it becomes clear that the text follows a well-organised structure – in fact almost as meticulously crafted as that of *Philosophical Investigations* – which has not been taken into account by previous commentators. Wittgenstein composed the text of Appendix III as a thorough and systematic discussion, at the centre of which figures the following kind of question: how are we supposed to react to statements such as the one presented in section 8 of the appendix, *viz*. '"I have constructed a sentence (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'.…."' (Appendix III,

§8)? As we saw, Wittgenstein intended to ask in particular: what might be the implications of such a statement for the *mathematical practice* that it purports to be addressing?

Contrary to the assumption of most commentators, it is therefore not Gödel's proof as such that is central to the dialectic of Appendix III. Wittgenstein deliberately bypasses any of the specifics of Gödel's technical procedure, and he charitably assumes the most favourable conditions of complete formal correctness. Instead, Wittgenstein's focus is exclusively on Gödelian *explanations*, not of the internal correctness of the proof, but of the proof's alleged position within the existing system of mathematics.

More specifically, in Appendix III, Wittgenstein undertakes a thorough examination of what *might* be the exact meaning of some of the most sensational and apparently significant explanations of Gödel's results (namely, those that originated from Gödel's own original explanations in 1930/31). And he takes great care indeed to avoid the appearance of trying to do anything over and above his, as such, well-defined aims.

The outcome of Wittgenstein's examination is, in one important sense, very Socratic. He does not succeed in determining the exact meaning of those explanations. Ultimately, it remains evidently opaque what the mathematical role of *P* might possibly be, or if indeed it was ever supposed to have such a role. And, insofar as this is the case, Gödel was exactly right to think that, as he writes to Menger, 'Wittgenstein did not understand it' (2003 [1972], p.133). And Gödel was equally right to think that, in Appendix III, Wittgenstein also 'pretended not to understand it' (*ibid*.). However, it appears that Gödel did not entirely grasp *why* this was so. — The reason, namely, is that – by way of thorough philosophical investigation of the kind we are asked to undertake in Appendix III – Wittgenstein, unlike Gödel, had come to *understand* that he did not understand and also, to some extent, what some of the general obstacles are to an understanding of Gödel's results such as Wittgenstein was aspiring to (i.e. philosophical clarity). Thus, while Gödel indeed showed some significant understanding of Wittgenstein here, ultimately, Wittgenstein perhaps understood Gödel better than Gödel understood himself.[73]

[73] This chapter was originally intended to be a translation of a shortened version of Kienzler (2008). Over time it has evolved into a piece in its own right.

Postscript

In the final proofs of his 1931 paper, Gödel introduced a change of phrasing on the first page. Originally he had written the following (referring to the systems of *PM* and *Zermelo-Fraenkel*):

> One might therefore conjecture that these axioms and rules of inference are sufficient to **carry out any conceivable proof** [*überhaupt jeden denkbaren Beweis zu führen*].[74]

He altered this to read:

> One might therefore conjecture that these axioms and rules of inference are sufficient to **decide *any* mathematical question that can at all be formally expressed in these systems**.

While this change specifies the scope of his proof more accurately, by mentioning the relevant systems, it also moves away from the more natural notion of *carrying out proofs*. Instead, Gödel now expressly claims that he is dealing with *mathematical questions* – questions of course being a kind of sentence – that are *formally expressed* and (possibly) *decided* (as it were, in two distinct mathematical procedures).

This may indicate that Wittgenstein's questioning of the assimilation of proofs to sentences comes very close to the heart of Gödel's enterprise.

References

Bays, T. (2004) 'On Floyd and Putnam on Wittgenstein and Gödel', *Journal of Philosophy* 101, pp.197–210.

Berto, F. (2009) 'The Gödel Paradox and Wittgenstein's Reasons', *Philosophia Mathematica* 17:2, pp.208–19.

Feferman, S. (1984) 'Kurt Gödel: Conviction and Caution', *Philosophia Naturalis* 21:2/4, pp.546–62.

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: Essays on the Development of the Foundations of Mathematics*, Kluwer, pp.373–425.

[74] The first page of these proofs, including Gödel's changes, is reproduced in Sigmund, Dawson and Mühlberger, 2006, p.114. The translation of the crossed-out phrase is ours.

Floyd, J. (2001) 'Prose versus Proof: Wittgenstein on Gödel, Tarski, and Truth', *Philosophia Mathematica* 9:3, pp.280–307.

Floyd, J. (2012) '*Das Überraschende*: Wittgenstein on the Surprising in Mathematics', in J. Ellis and D. Guevara (eds) *Wittgenstein and the Philosophy of Mind*, Oxford University Press, pp.225–58.

Floyd, J. and Putnam, H. (2000) 'A Note on Wittgenstein's Notorious Paragraph about the Gödel Theorem', *Journal of Philosophy* 97, pp.624–32.

Frege, G. (1979) *Posthumous Writings*, H. Hermes, F. Kambartel and F. Kaulbach (eds), trans. P. Long and R. White, Blackwell.

Gödel, K. (1930) 'Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit', *Anzeiger der Akademie der Wissenschaften in Wien* 61, pp.214–15. English translation in Gödel (1986).

Gödel, K. (1931) 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38, pp.173–98. English translation in Gödel (1986).

Gödel, K. (1934) 'On Undecidable Propositions of Formal Mathematical Systems' (mimeographed lecture notes, taken by S. Kleene and J. Rosser), reprinted with corrections in Gödel (1986).

Gödel, K. (1986) *Collected Works. Volume I. Publications 1929–1936*, S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay and J. van Heijenoort (eds), Oxford University Press.

Gödel, K. (2003) *Collected Works. Volume V. Correspondence H–Z*, S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons and W. Sieg (eds), Oxford University Press.

Kienzler, W. (2008) 'Wittgensteins Anmerkungen zu Gödel. Eine Lektüre der *Bemerkungen über die Grundlagen der Mathematik*, Teil I, Anhang III', in M. Kroß (ed) *'Ein Netz von Normen': Wittgenstein und die Mathematik*, Parerga, pp.149–98.

Mühlhölzer, F. (2002) 'Wittgenstein and Surprises in Mathematics', in R. Haller and K. Puhl (eds) *Wittgenstein and the Future of Philosophy. A Reassessment after 50 Years,* Österreichischer Bundesverlag, pp.306–15.

Rodych, V. (1999) 'Wittgenstein's Inversion of Gödel's Theorem', *Erkenntnis* 51, pp.173–206.

Sigmund, K., Dawson, J. and Mühlberger, K. (eds) (2006) *Kurt Gödel. Das Album / The Album*, Vieweg.

Shanker, S. G. (1988) 'Wittgenstein's Remarks on the Significance of Gödel's Theorem', in S. G. Shanker (ed) *Gödel's Theorem in Focus*, Croom Helm, pp.155–256.

Steiner, M. (2001) 'Wittgenstein as His Own Worst Enemy: The Case of Gödel's Theorem', *Philosophia Mathematica* 3, pp.257–79.

Whitehead, A. N. and Russell, B. (1910–13, ^{2}1925–27) *Principia Mathematica*, Cambridge University Press.