Wittgenstein's "notorious paragraph" about the Gödel Theorem

Timm Lampert, University of Bern, Switzerland

1. Introduction

In §8 of *Remarks on the Foundations of Mathematics* (RFM), Appendix 3 Wittgenstein imagines what conclusions would have to be drawn if the Gödel formula P or ¬P would be derivable in PM. In this case, he says, one has to conclude that the interpretation of P as "P is unprovable" must be given up. This "notorious paragraph" has heated up a debate on whether the point Wittgenstein has to make is one of "great philosophical interest" revealing "remarkable insight" in Gödel's proof, as Floyd and Putnam suggest (Floyd (2000), Floyd (2001)), or whether this remark reveals Wittgenstein's misunderstanding of Gödel's proof as Rodych and Steiner argued for recently (Rodych (1999, 2002, 2003), Steiner (2001)). In the following the arguments of both interpretations will be sketched and some deficiencies will be identified. Afterwards a detailed reconstruction of Wittgenstein's argument will be offered. It will be seen that Wittgenstein's argumentation is meant to be a rejection of Gödel's proof but that it cannot satisfy this pretension.

The notorious paragraph runs as follows (the last three sentences are omitted and will not be discussed in this paper):

> I imagine someone asking my advice; he says: "I have constructed a proposition (I will use 'P' to designate it) in Russell's symbolism, and by means of certain definitions and transformations it can be so interpreted that it says 'P is not provable in Russell's system'. Must I not say that this proposition on the one hand is true, and on the other hand is unprovable? For suppose it were false; then it is true that it is provable. And that surely cannot be! And if it is proved, then it is proved that is not provable. Thus it can only be true, but unprovable."
>
> Just as we ask, "'Provable' in what system?", so we must also ask, "'true' in what system?" 'True in Russell's system' means, as was said: proved in Russell's system; and 'false in Russell's system' means: the opposite has been proved in Russell's system. – Now what does your "suppose it is false" mean? *In the Russell sense* it means 'suppose the opposite is proved in Russell's system'; *if that is your assumption* you will now presumably give up the interpretation that it is unprovable. And by 'this interpretation' I understand the translation into this English sentence. – If you assume that the proposition is provable in Russell's system, that means it is true *in the Russell sense*, and the interpretation "P is not provable" again has to be given up.[…]

2. The Floyd-Putnam interpretation: Wittgenstein's "remarkable insight"

While the first commentators such as Kreisel (1958) and Bernays (1959) were rather shocked by Wittgenstein's remarks on Gödel in the 1956 published RFM and concluded that Wittgenstein failed to appreciate the consistency assumption of Gödel's proof, Floyd argued in her papers 2000 (with Putnam) and 2001 contrary to Kreisel and Bernays that Wittgenstein's argumentation is based on Gödel's assumption of ω-consistency. Relying on the above quoted §8 Floyd and Putnam attribute to Wittgenstein a "philosophical claim of great interest", namely that "if one assumes […] that ¬P is provable in Russell's system one should […] give up the 'translation' of P by the English sentence 'P is not provable'" (Floyd (2000), p. 625). According to Floyd and Putnam this claim is grounded in Wittgenstein's acceptance of Gödel's mathematical proof showing that PM must be ω-inconsistent, if ¬P is provable. From this it follows that the predicates 'NaturalNo(x)' and 'Proof(x,t)' occurring in P cannot be interpreted as 'x is a natural number' and 'x is the number of a proof of the formula with the number t' because one has to allow for non-standard interpretations of the variable's values being other than numbers. This interpretation is joined by the claim that Wittgenstein's "aim is not to refute the Gödel theorem", "for nothing in that proof turns on any such translation into ordinary language" (Floyd (2000), p. 625f). According to Floyd and Putnam Wittgenstein himself is just stating what Gödel holds if the latter insists that his proof is independent of any interpretation and rests on consistency assumptions.

Depart from the question of the "philosophical interest" of the position Floyd and Putnam attribute to Wittgenstein, the sympathetic point of their interpretation seemed to be the possibility to find a way of interpreting Wittgenstein's remarks on Gödel without the need of accusing him to misunderstand the great mathematician. However, depart from the fact that they cannot quote direct textual evidence for their interpretation, two main reasons speak against it: First of all, it cannot be ignored that Wittgenstein does not agree with Gödel's argument, to mention only some evidence: he calls the "Gödelian reason" a "stupid one", that is "obviously nonsensical" (MS121, 81v), similar to paradoxes he sees in Gödel's argument a "profitless performance" (RFM, App.3, §12) and "bits of legerdemain" (RFM, App.3, §19), he questions to call Gödel's proof a "forcible reason" for giving up the search of a proof (RFM, App.3, §14f.). Secondly, what Wittgenstein says in §8 is not in harmony with Gödel's argument. Wittgenstein presumes that "P is true in Russell's sense" means "P is provable in Russell's system" and "P is false in Russell's sense" means "¬P is provable in Russell's system" – these assumptions Gödel would not affirm but claim to have disproved by his incompleteness proof. And he (Wittgenstein) maintains that due to the derivation of a contraction from both of the assumptions – that ¬P is provable and that P is provable – one will follow that the interpretation of P as "P is unprovable" has to be given up. Yet, Gödel in his proof shows that assuming the provability of P it would follow that PM is ω-inconsistent and assuming P it would follow that PM is inconsistent and from this he draws the conclusion that these assumptions and not any interpretation have to be given up. Whatever the exact meaning of Wittgenstein's argumentation in §8 is, to try to harmony it with Gödel's view does not justice to both – Wittgenstein and Gödel. Actually, Gödel rejected Wittgenstein's argumentation accusing him to confuse his argumentation with "a kind of logical paradox" (cf. Wang (1987), p. 49).

According to Floyd and Putnam, Wittgenstein's "remarkable insight" culminates in "a philosophical claim of great interest", namely that Gödel does not prove the truth of P, because this would presuppose the acceptance of the translation of P which is in question because the consistency of Russell’s system PM cannot be proven. Even Rodych (1999), p. 188ff. makes a similar claim according to his interpretation of §14f.and RFM, VII 22. Unfortunately, Floyd, Putnam and Rodych do not explain why the triviality that Gödel’s proof including the claim that P is true rests on the assumption of consistency of PM is a “philosophical claim of great interest”. They leave it an open question what the significant difference consists in whether one says that Gödel proved the undecidability and truth of P, given the consistency of PM, or one insists that Gödel did not prove the truth of P because the consistency of PM is not proven. Thus, if one does not trace Wittgenstein’s remark on Gödel back to a misunderstanding all what one can make out of them is a repetition of the fact that Gödel’s proof rests on the unproven assumption of consistency as Gödel stressed himself.

3. The Rodych-Steiner interpretation: Wittgenstein’s mistake

According to Rodych and Steiner §8 cannot be interpreted without accusing Wittgenstein to misunderstand Gödel’s argumentation. To both of them Wittgenstein mistakenly assumes that Gödel’s proof rests on a natural language interpretation, whereas the pure mathematical part proving the undecidability of either P or $\neg$P does not presume any interpretation of P. As will be seen, evidence can be put forward for this claim. Yet, whereas Steiner does not at all argue for it by examining §8, Rodych (1999) does offer a detailed analysis of §8, yet not a wholly convincing one. In Rodych (1999), p.182 Wittgenstein’s mistake is explained in the following way:

> Thus, when Wittgenstein says in §8 “*if that* [i.e. “‘suppose ‘P’ is false’ means ‘suppose the opposite is proved in Russell’s system’”] *is your assumption*”, the obvious and immediate Gödelian reply is: “Well, yes I would ‘now presumably give up the interpretation that it is unprovable’ if that were *my* assumption – but it *isn’t* – it’s *your* assumption”.

The parenthesis is by Rodych and interprets the reference of “that” wrongly: The demonstrative pronoun does not refer to Wittgenstein’s interpretation of “P is false” in the sense of “$\neg$P is proved in Russell’s system”, but only to the assumption “$\neg$P is proved in Russell’s system” – this is the assumption directly mentioned before. Wittgenstein maintains in §8 that this assumption will compel one to give up the interpretation of P as “P is unprovable”. To this, the obvious and immediate Gödelian reply would not be, as Rodych maintains, that one would give up the interpretation of P as “P is unprovable”, whether this assumption is only Wittgenstein’s or not. The immediate Gödelian reply would rather be: Given this assumption, it follows by applying purely recursive definitions, an ω-inconsistency, and that is why I actually do give up this assumption! Thus, a Gödelian needs not to accept Wittgenstein’s argumentation and draw the consequence that P shall not be interpreted as “P is unprovable”. Though not needed in order to derive the undecidability of P in PM, as will be seen in the next section, this interpretation is needed in order to prove the incompleteness of P. Thus from a Gödelian point of view one should not concede Wittgenstein’s argumentation and assume the provability of $\neg$P while giving up the interpretation of P as “P is unprovable”.

To sum up, in order to evaluate the validity of Wittgenstein’s argument a thoroughly reconstruction of it and a comparison to Gödel’s way of argumentation is needed. This will be done in the following section, by examining sentence for sentence of the above quoted §8.

4. Reconstruction of Wittgenstein’s argument

§8 opens by presuming that “by means of certain definitions and transformations” P can be interpreted by “P is not provable in Russell’s system”. Let’s abbreviate this assumption by

$$P = \neg\Pi\ P$$

In the following of the first paragraph an argument is given for the thesis that P is true and unprovable by reducing the negation of both conjuncts of this thesis to absurdity. First of all from $\neg$P a contradiction is derived (sentence 3 and 4). In order to do so, in addition to the assumption $P = \neg\Pi\ P$ the assumption $\Pi\ A \rightarrow A$, i.e. the correctness of PM, has to be introduced. The reductio of $\neg$P can be reconstructed in the following way:

1*	(1)	$\neg P$	A
2	(2)	$P = \neg\Pi P$	A
1*,2	(3)	$\neg\neg\Pi P$	2,1=E
1*,2	(4)	ΠP	3 DNE
5	(5)	$\Pi A \rightarrow A$	A
5	(6)	$\Pi P \rightarrow P$	5 SUB
1*,2,5	(7)	P	6,4 MPP
1*,2,5	(8)	$P \& \neg P$	7,1 &I
2,5	(9)	$\neg\neg P$	1,8 RAA
2,5	(10)	P	9 DNE

In the last but one sentence of paragraph 1 Wittgenstein hints at the reductio of the second conjunct of the thesis that P is true and unprovable (i.e. $P \& \neg\Pi P$). In detail, the argumentation runs as follows:

1*	(1)	ΠP	A
2	(2)	$P = \neg\Pi P$	A
1*,2	(3)	$\Pi\neg\Pi P$	2,1=E
4	(4)	$\Pi A \rightarrow A$	A
4	(5)	$\Pi\neg\Pi\ P\ \ \rightarrow \neg\Pi P$	4 SUB
1*,2,4	(6)	$\neg\Pi P$	5,3 MPP
1*,2,4	(7)	$\Pi P \& \neg\Pi P$	1,6&I
2,4	(8)	$\neg\Pi P$	1,7 RAA

The last sentence of the first paragraph of §8 entails the conclusio of this argumentation, i.e. the conjunction of both arguments: $P \& \neg\Pi P$.

A similar, though not identical proof sketch is given by Wittgenstein in MS 117, pp.147-148. Of course, this way of understanding Gödel’s proof, is mistaken. First of all, Gödel never starts by assuming the truth or falsity of P, even not in his introductory remarks of Gödel (1931).

Instead, Gödel always gives a reductio of the provability of P and the provability of ¬P. Secondly, his reductio, as conveyed in his formal proof following his introduction does not entail the “interpretation assumption” P = ¬Π P. Instead, he is relying on purely recursive definitions (in short: DEF.). Presupposing Rosser’s improvement of Gödel’s proof that allows one to dispense with the concept of ω-inconsistency these definitions allow one to *construct* a proof for P given a proof for ¬P and vice versa. The “interpretation assumption” is only needed in order to conclude from the undecidability of P (i.e. from ¬ΠP & ¬Π¬P) the incompleteness of PM (i.e. P & ¬ΠP). Thus, in order to compare Gödel’s way of arguing, his proof can be put in the following form: Given ΠP one yields only by applying DEF. a proof of ¬P, ergo Π¬P, ergo ΠP & Π¬P, presuming the consistency of PM, it follows by RAA ¬ΠP; given Π¬P, it follows only by applying DEF. a proof of P, ergo ΠP, ergo ΠP & Π¬P, presuming the consistency of PM, it follows by RAA ¬Π¬P. Ergo ¬ΠP & ¬Π¬P (the “undecidability thesis”). Yet, given P = ¬ΠP (the “interpretation assumption”), P follows form the first conjunct of the undecidability thesis. Thus, Gödel’s proofs of the undecidability thesis is a mathematical proof in the sense that it is only relying on recursive definitions, yet his proof of the incompleteness of PM is based on the interpretation assumption. Because he wants to end up with the incompleteness of PM, a Gödelian is not willing to give up this assumption and even though he will not accept Wittgenstein’s reconstruction of the argument, he also will not be inclined to draw the consequence Wittgenstein wants him to draw in the second paragraph of §8, that now shall be examined further.

Wittgenstein’s argument starts with the assumption, he elaborated in the §1-7: “truth” is - as “provability” - system-dependent; “true in Russell’s system” means “proved in Russell’s system”, “false in Russell’s system” means “the opposite has been proven in Russell’s system. Thus, one yields

P = ΠP

¬P = Π¬P

Given this, he argues in the following, his opponent will not any more be inclined to draw the same consequences: Instead of reducing ¬P and ΠP to absurdity, his opponent will now come to understand that the “interpretation assumption” P = ¬ΠP has to be given up, because if he would still derive the thesis P & ¬ΠP this would be contradictory: According to ¬P = Π¬P this would amount to P & ¬P and according to P = ΠP this would amount to maintain ΠP & ¬ΠP. Thus, Wittgenstein’s argument is, that the reductio argumentation of his opponent are underdetermined and that by considering P = ΠP and ¬P = Π¬P the Gödelian will come to understand that not the assumption ¬P and Π¬P have to be reduced to absurdity but the interpretation assumption P = ¬ΠP.

This reasoning is mistaken because of the following reasons:

1. Gödel does not agree with the assumptions Wittgenstein starts his argumentation in the second paragraph: Whether P = ΠP and ¬P = Π¬P are valid is just what is in question and the philosophical upshot of Gödel’s proof is to have *proven* that these assumptions are wrong. This, indeed, is in conflict with Wittgenstein’s philosophy. Yet, Wittgenstein has to argue against the premises of Gödel’s proof (especially DEF. which itself strengthen the interpretation assumption), if he wants to stick to these assumptions. One cannot pertain to argue against the incompleteness proof by presupposing the falsehood of its conclusion.
2. Gödel does not start his argument, by presuming ¬P and reducing it to absurdity: Instead, he only reduces Π¬P and ΠP to absurdity, thus putting forward the undecidability thesis ¬Π¬P & ¬ΠP. And this he does without assuming the interpretation assumption. Only his move from the undecidability thesis towards the incompleteness theorem presupposes the interpretation assumption without hereby using RAA.

Conclusion

According to any given interpretation, Wittgenstein’s notorious remark on Gödel cannot be appreciated as revealing a “remarkable insight” of “great philosophical interest”, because either it is understood as simply affirming what Gödel said or as a misguided critique of Gödel’s proof. Wittgenstein’s argumentation is no challenge for the Gödelian, yet Gödel’s argumentation is a challenge for the Wittgensteinian.

References

Bernays, Paul 1959, “Comments on Ludwig Wittgenstein’s *Remarks on the Foundations of Mathematics*”, *Ratio 2.1*, 1-22.

Floyd, Juliet and Putnam, Hilary 2000, “A Note on Wittgenstein’s ‘Notorious Paragraph’ about the Gödel Theorem”, *The Journal of Philosophy* XCVII,11, 624-632.

Floyd, Juliet 2001, “Prose versus Proof: Wittgenstein on Gödel, Tarski, and Truth”, *Philosophia Mathematica* 3.9, 280-307.

Gödel, Kurt 1931, “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme”, *Monatshefte für Mathematik und Physik 38*, 173-198.

Kreisel, Georg 1958, “Wittgenstein’s Remarks on the Foundations of Mathematics”, *British Journal for the Philosophy of Science 9*, 135-57.

Rodych, Victor 1999, “Wittgenstein’s Inversion of Gödel’s Theorem”, *Erkenntnis* 51, 173-206.

Rodych, Victor 2002, “Wittgenstein on Gödel: The Newly Published Remarks”, *Erkenntnis* 56, 379-397.

Rodych, Victor 2003, “Misunderstanding Gödel: New Arguments about Wittgenstein and New Remarks by Wittgenstein”, *Dialectica* 57, 279-313.

Steiner, Mark 2001, “Wittgenstein as His Own Worst Enemy: The Case of Gödel’s Theorem”, *Philosophica Matehmatica 9*, 257-279.

Wang, Hao 1987, *Reflections on Kurt Gödel*, Cambridge: MIT Press.

Wittgenstein, Ludwig *1978, Remarks on the Foundations of Mathematics* (RFM), Oxford: Blackwell.

Wittgenstein, Ludwig 2000, *Wittgenstein’s Nachlass: The Bergen Electronic Edition*, Oxford: University Press.