Wittgenstein and Gödel: An Attempt to Make "Wittgenstein's Objection" Reasonable*

Timm Lampert

published in *Philosophia Mathematica* Philosophia Mathematica 25.3, 2018, p. 324-345

Abstract

According to some scholars, such as Rodych and Steiner, Wittgenstein objects to Gödel's undecidability proof of his formula G, arguing that given a proof of G, one could relinquish the meta-mathematical interpretation of G instead of relinquishing the assumption that *Principia Mathematica* (PM) is correct (or ω-consistent). Most scholars agree that such an objection, be it Wittgenstein's or not, rests on an inadequate understanding of Gödel's proof. In this paper, I argue that there is a possible reading of such an objection that is, in fact, reasonable and related to Gödel's proof.

1 Introduction

Wittgenstein's discussion of Gödel's incompleteness proof[1] has elicited various reactions. Whereas Wittgenstein was accused by his earliest critics of having misunderstood Gödel[2], recent interpreters have attempted to do better justice to Wittgenstein by separating the wheat from the chaff in his remarks.

Overall, recent interpretations are characterized by the tendency to draw value from Wittgenstein's remarks by separating them from Gödel's undecidability proof. By contrast, I will argue that Wittgenstein's critique can be more closely related to Gödel's undecidability proof to explain why this particular proof need not necessarily lead one to give up the search for decision procedures concerning provability within *Principia Mathematica* (PM).

*Parts of this paper were presented in different colloquia at the Humboldt University Berlin. I thank the audiences there for discussions. I am also grateful to two anonymous referees of *Philosophia Mathematica* for their helpful comments.

[1] Cf., in particular, RFM, part I, appendix I,§1-19 (I,§1-19 for short), and RFM, part VII,§21-22 (VII,§21-22 for short). Rodych [2002] compiles further remarks on Gödel from Wittgenstein's Nachlaß, published by the Bergen project.

[2] Cf. Gödel himself in Wang [1987, p. 49] as well as Anderson [1958], Kreisel [1958], Bernays [1959], and Dummett [1959].

Contrary to Wittgenstein's early critics, Shanker [1988], Floyd & Putnam [2000] and Floyd [2001] argue that Wittgenstein does not question Gödel's undecidability proof itself. Instead, they say, Wittgenstein's remarks are concerned with the semantic and philosophical consequences of Gödel's proof; those remarks represent, according to Floyd and Putnam, a "remarkable insight"[3] regarding Gödel's proof. I share the view that Wittgenstein believed that it is not the task of philosophy to question mathematical proofs. However, I argue that from Wittgenstein's perspective, Gödel's proof is not a mathematical proof. Instead, it is a proof that relies on "prose" in the sense of meta-mathematical interpretations, and thus, it is a valid object of philosophical critique. Thus, I deny that Wittgenstein views Gödel's undecidability proof as being just as conclusive as mathematical impossibility proofs. Wittgenstein's simplified, rather general way of referring to an ordinary language interpretation of G without specifying exactly where questionable meta-mathematical interpretations are relevant to Gödel's proof might have led to the judgment that Wittgenstein's critique is not relevant to Gödel's proof.

Contrary to Floyd and Putnam, Rodych [1999] and Steiner [2001] assume that Wittgenstein argues against Gödel's undecidability proof. According to their interpretation, Wittgenstein's objection against Gödel's proof is that from proving G or $\neg G$, it does not follow that PM is inconsistent or ω-inconsistent.[4] Instead, one could abandon the meta-mathematical interpretation of G. However, according to both authors, this critique is inadequate because Gödel's proof does not rely on a meta-mathematical interpretation of G. By specifying where Wittgenstein's critique is mistaken, they wish to decouple Wittgenstein's philosophical insights from his mistaken analysis of Gödel's mathematical proof. I agree with Rodych and Steiner that Wittgenstein's critique does not offer a sufficient analysis of the specific manner in which a meta-mathematical interpretation is involved in Gödel's reasoning. However, in contrast to these authors, I will explain why both Gödel's semantic proof and his so-called syntactic proof do rely on a meta-mathematical interpretation.

Priest [2004], Berto [2009a] and Berto [2009b] view Wittgenstein as a pioneer of paraconsistent logic. They are especially interested in Wittgenstein's analysis of Gödel's proof as a proof by contradiction. Like Rodych and Steiner, they maintain that Wittgenstein's remarks are not, in fact, pertinent to Gödel's undecidability proof because Wittgenstein refers not to a syntactic contradiction within PM but rather to a contradiction between the provability of G and its meta-mathematical interpretation. However, according to them, Wittgenstein's critique is not mistaken. Rather, it is concerned with the interpretation and consequences of Gödel's undecidability proof. Presuming Wittgenstein's rejection of any distinction between (i) metalanguage and object language and (ii) provability and truth, they show that engaging with Gödel's proof depends

[3]Floyd & Putnam [2000, p. 627], in accordance with Goodstein [1957, p. 551]; cf. Bays [2006] for a critique of Floyd & Putnam [2000] and Floyd & Putnam [2006] for a reply.

[4]I refer to PM, and not to *Peano Arithmetic* (PA), as is usual at present. However, for the argument presented in this paper, this difference is not relevant. By "proof" or "$\vdash$", I refer to a proof within PM, unless otherwise indicated.

on philosophical presumptions. I do not question this. However, I will argue that Wittgenstein's critique can be interpreted in a way that is indeed relevant to Gödel's undecidability proof.

The intention of this paper is not to enter into an exegetical debate on whether Wittgenstein understands Gödel's proof and whether he indeed objects to it. For the sake of argument, I assume that to be given. Furthermore, similarly to, e.g., Rodych and Steiner, I take "Wittgenstein's objection" to Gödel's proof to be as follows: "Instead of inferring the incorrectness or (ω-)inconsistency of PM (or PA) from a proof of G (or $\neg G$), one might just as validly abandon the meta-mathematical interpretation of G. Therefore, Gödel's proof is not compelling because it rests on a doubtful meta-mathematical interpretation." I recognize that this is highly controversial, to say the least. However, the literature seems to agree that such an objection, be it Wittgenstein's or not, has no relation to Gödel's undecidability proof[5] and thus is not reasonable. The intention of this paper is to show that this is not quite true. This objection can, indeed, be related to Gödel's method of defining provability within the language of PM, and it questions this essential element of Gödel's meta-mathematical proof method by measuring its reliability on the basis of an algorithmic conception of proof.

To show this, I first distinguish between algorithmic and meta-mathematical proofs (sections 2 and 3) and then relate Wittgenstein's objection to Gödel's semantic and syntactic versions of his proof against this background (section 4). In doing so, I pass over many controversies in the literature, adopt several simplifications and do not do the necessary exegetical work to argue in detail why I claim that Wittgenstein's non-revisionist understanding of mathematics applies to algorithmic proofs but not to meta-mathematical proofs involving meta-mathematical interpretations. This all serves the systematic purpose of explaining the sense in which Wittgenstein's objection is reasonable. Thus, although I often, for the sake of brevity, ascribe directly to Wittgenstein views that I advance in support of the reading of his objection that I propose in this paper, the reader may better understand my argument as a Wittgenstein-inspired attempt to explain "Wittgenstein's objection". By providing such an explanation, I hope to stimulate a debate on (i) whether this argument is indeed Wittgenstein's and (ii) whether it is indeed reasonable.

2 Algorithmic Proof

In I,§17, Wittgenstein suggests to look at proofs of unprovability "in order to see *what* has been proved". To this end, he distinguishes two types of proofs of unprovability. He mentions the first type only briefly: "Perhaps it has here been proved that such-and-such forms of proof do not lead to P." (P is Wittgenstein's abbreviation for Gödel's formula G). In this section, I argue that Wittgenstein refers in this quote to an algorithmic proof proving that G is not provable within

[5] Cf., e.g., [Rodych, 1999, pp. 182f.], quoted on p. 14 and in footnote 17 on p. 18; [Steiner, 2001, p. 259, p. 263]; and [Floyd & Putnam, 2000, p. 625].

PM.[6] Such a proof of unprovability would, to Wittgenstein, be a compelling reason to give up the search for a proof of G within PM. Wittgenstein challenges Gödel's proof because it is not an unprovability proof of this type. This is also why Wittgenstein does not consider algorithmic proofs of unprovability in greater detail in his discussion of Gödel's proof. Such proofs represent the background against which he contrasts Gödel's proof to a type of proof that is beyond question.

Unfortunately, Wittgenstein does not follow his own suggestion to more carefully evaluate unprovability proofs with respect to Gödel's proof. Instead, he distinguishes different types of proofs of unprovability in his own words and in a rather general way; cf. I,§8-19. His critique focuses on a proof of unprovability that relies on the representation of provability within the language of the axiom system in question. Thus, following his initial acknowledgment of algorithmic unprovability proofs in I,§17, Wittgenstein repeats, at rather great length, his critique of a meta-mathematical unprovability proof. It is this type of unprovability proof that he judges unable to provide a compelling reason to give up the search for a proof of G. The most crucial aspect of any comparison of two different types of unprovability proofs is the question of what serves as the "criterion of unprovability" (I,§15). According to Wittgenstein, such a criterion should be a purely syntactic criterion independent of any meta-mathematical interpretation of formulas. It is algorithmic proofs relying on nothing but syntactic criteria that serve as a measure for assessing meta-mathematical interpretations, not vice versa.

Wittgenstein distinguishes Gödel's undecidability proof from algorithmic proofs of unprovability in I,§14. Here, he admits the validity only of algorithmic unprovability proofs that provide a compelling reason to give up the search for a proof and finally notes that an unprovability proof in the form of a proof by contradiction does not satisfy this condition.[7] As examples of acceptable impossibility proofs, Wittgenstein refers to classical geometric impossibility proofs proving that certain geometric problems are not solvable within Euclidean geometry. He explicitly mentions the trisection of angles with a ruler and compass; elsewhere, he also refers to the construction of polygons (cf., e.g., WVC, p. 36f.; PR, §152; and LFM, p. 56). According to Wittgenstein, questions regarding the possibility of certain constructions with a ruler and compass do not receive precise meaning until said constructions are transformed into algebraic expressions. As soon as this is done, those questions become decidable: those and only those angles that are expressible in terms of nested square roots are possible to construct. This sort of proof rests on nothing but a mathematical equivalence transformation for the purpose of deciding the initial question regarding the

[6]Like Wittgenstein in the above quote and in most of his remarks, I focus on the (un)provability of G rather than on that of $\neg G$.

[7]As Rodych [1999, p. 192] (cf. also Matthíasson [2013, p. 45]) accurately notes, Floyd [2001] misses that geometric impossibility proofs such as that for the trisection of arbitrary angles are paradigmatic of the algorithmic proofs with which Wittgenstein contrasts Gödel's meta-mathematical proof. Steiner [2001, p. 262-263] also criticizes Floyd's interpretation of §14. However, he shares her view that Gödel's proof – contrary to Wittgenstein's understanding – is similar to geometric impossibility proofs; cf. Steiner [2001, p. 258].

distinction between possible and impossible constructions of angles by applying a purely syntactic criterion that refers to the form of the resulting expressions. According to Wittgenstein, geometric impossibility proofs are not based on a problematic kind of proof by contradiction. Instead, they are paradigmatic of algorithmic proofs.[8]

Geometric impossibility proofs reduce the question of the possibility of geometric constructions to the question of whether certain algebraic equations are solvable using numbers of a certain algebraic form. Impossibility proofs proving that certain equations are not solvable using numbers of a certain form are additional prominent examples of impossibility proofs that are acceptable according to Wittgenstein. As a demanding example, one might consider Galois theory deciding whether an algebraic equation has solutions in terms of radicals. Wittgenstein mentions simpler examples. One is the application of the Euclidean division algorithm to determine the possibility of representing a rational number using a finite decimal number. This algorithm yields either a finite representation as a result, e.g., 0.25 in the case of $\frac{1}{4}$, or a repeated finite computation. The observation of such repetition can serve as a syntactic criterion for identifying the impossibility of a finite representation, e.g., the repetition of dividing 10 by 3 when applying the Euclidean algorithm to $\frac{1}{3}$ (cf. WVC, p. 33 and 135; PR, §187; and PG, appendix, §35). Likewise, the proof of the irrationality of $\sqrt{2}$ is not a proof of contradiction according to Wittgenstein. Instead, it is an example of a proof "by induction" that relates to the repetition of a computation (here, in solving the equation $x^2 = 2$ within the system of rational numbers) as a syntactic criterion of impossibility; cf. WVC, p. 136. The same holds for Euclid's proof of the impossibility of enumerating all primes within a finite number of steps; cf. ibid. Again, the proof consists of providing a rule that permits the generation of another prime through repeated application. A traditional reformulation of these proofs in terms of informal proofs by contradiction is not a sufficient reason to classify those proofs as proofs by contradiction according to Wittgenstein's understanding. As is explained below on p. 6, such informal proofs, in Wittgenstein's view, reduce meta-mathematical statements, such as "$x^2 = 2$ is solvable within the rational numbers", to absurdity by *presuming algorithmic proofs*. Unprovability proofs, or impossibility proofs in general, in the form of algorithmic proofs prove that certain problems are not solvable according to *given* rules applied to certain forms of expressions. In the context of algorithmic unprovability proofs, the "criteria of unprovability" are purely syntactic decision criteria. Such a proof terminates with an expression comprising syntactic properties to offer a decision on provability.

In propositional logic, an algorithmic conception of proof is realized, e.g., by transforming the negation of a formula into its disjunctive normal form through a logical equivalence transformation and deciding whether each disjunct contains an atomic formula and its negation. If this is not the case, then the initial

[8] Mühlhölzer [2001], footnote 16, characterizes geometric impossibility proofs as proofs by contradiction and refers to Martin [1998, pp. 28,45]. Such a reconstruction, however, is misleading with regard to the conception of Wittgenstein's proof, and in fact, it is not supported by the cited reference, either.

formula is not provable. Likewise, Venn diagrams or tree calculi allow one to decide upon provability within monadic first-order logic. From a Wittgensteinian perspective, a relevant question in the case of polyadic first-order logic is the manner in which unprovability proofs are possible (i) by transforming first-order formulas into equivalent expressions that allow one to decide on the provability of the initial formulas or (ii) by identifying repeated applications of derivation rules within a suitable calculus. Proofs of unprovability within a logic based on decision procedures that do not refer to interpretations of formulas are acceptable according to Wittgenstein. If it were possible to prove in this way that neither G nor $\neg G$ follows logically from the axioms of PM, then Wittgenstein would not question the undecidability of PM. It is this type of unprovability proof to which Wittgenstein refers in I,§17 by saying "that such-and-such forms of proof do not lead to P". Obviously, Gödel does not argue in this way. By contrast, by applying Gödel's method, Church proves that a decision procedure is impossible for the entire realm of first-order logic. Undecidability proofs such as those of Gödel and Church limit the range of algorithmic proofs, which challenges Wittgenstein's algorithmic proof conception. He answers this challenge by questioning the probative force of meta-mathematical undecidability proofs.

According to Wittgenstein, algorithmic proofs are paradigmatic of genuine mathematical proofs. His rejection of unprovability proofs in the form of proofs by contradiction does not imply rejection of proofs by contradiction in general. Indeed, for the case of mathematical algorithmic proofs, he emphasizes that statements that contradict the results of such proofs are reduced to absurdity (RFM VI,§28, first paragraph):

> We can always imagine proof by *reductio ad absurdum* used in argument with someone who puts forward a non-mathematical assertion (e.g. that he has seen a checkmate with such-and-such pieces) which can be mathematically refuted.

Here, as in other passages, Wittgenstein uses impossibility proofs from the domain of chess as paradigmatic algorithmic impossibility proofs; cf., e.g., the impossibility of reaching checkmate with a knight and a king (WVC, p. 133-136). Claims that are reduced to absurdity by algorithmic unprovability proofs are non-mathematical, or, more precisely, meta-mathematical, statements that contradict the results of said proofs. Although one might still question the calculus on which an algorithmic proof is based, such a proof serves as a measure for seeking a proof *within* the calculus (cf. WVC, p. 207f., which refers to the proof of the irrationality of $\sqrt{2}$; LFM, p. 47 and 56, which concerns geometric impossibility proofs; and RFM VI,§28, third paragraph, in general).

Gödel's proof is not an algorithmic unprovability proof. Instead, Gödel's proof is based on the representation of provability within the language of PM. Based on this assumption, Gödel concludes that PM would be inconsistent (or ω-inconsistent) if G (or $\neg G$) were provable. Thus, given PM's (ω)-consistency, G is undecidable. This reasoning is based on the purely hypothetical assumption of the provability of G; it does not consider any specific proof strategies for proving formulas of a certain form within PM.

Given an algorithmic unprovability proof for G, the meta-mathematical statement that G is provable would be reduced to absurdity. This would be a compelling reason to abandon any search for a proof. Such a proof by contradiction would contain a "physical element" (I,§14) because a *meta*-mathematical statement concerning the provability of G is reduced to absurdity on the basis of an algorithmic, and thus purely mathematical, proof. Wittgenstein does not reject such a proof by contradiction in §14. In the following section, I will argue that he instead rejects proofs by contradiction that concern the relation between the provability of a formula and its interpretation.

3 Proof by Contradiction

One might expect that Wittgenstein would define a proof by contradiction as a proof that derives a formula of the form $A \wedge \neg A$ or $a \neq a$ within a calculus. Applied to PM, this would prove the inconsistency of PM. Indeed, Wittgenstein does consider such a proof, deriving both G and $\neg G$ "directly", i.e., within PM, at the end of I,§17. However, such a proof is obviously not an unprovability proof. Instead, it is an algorithmic proof of $G \wedge \neg G$. Such a proof is unrelated to Gödel's proof because Gödel proves that neither G nor $\neg G$ is provable, presuming the (ω-)consistency of PM. Inconsistency proofs in the form of proving a formal contradiction within a calculus do not constitute a specific problem for Wittgenstein. He also does not object to proofs by contradiction in terms of reducing some meaningful proposition in a set of several meaningful propositions to absurdity; cf. RFM VI, §28, second paragraph. Such a proof simply proves that not all of the propositions are true.

The proofs by contradiction of the type to which Wittgenstein objects are proofs that involve the *interpretation* of logical formulas: the inconsistency concerns the relation between the provability of a formula (proven or merely assumed) and its interpretation. Here, "interpretation" is not to be understood in terms of purely formal semantics underlying proofs of correctness or completeness. Formal semantics assign extensions to formal expressions without considering specific instances of formal expressions that are meant to refer to extensions. Instead, in proofs of contradiction Wittgenstein is concerned with an "interpretation of a formula" refers to an *instance* of a formula or of its abbreviation, such as G or $\neg\exists y B(y, \ulcorner G \urcorner)$, stated as a sentence in ordinary language or a standardized fragment of ordinary language. Interpretations of this kind are so-called "intended interpretations" or "standard interpretations", which are intended to identify extensions such as truth values, truth functions, sets or numbers by means of ordinary expressions. As soon as interpretations of this kind become involved, one departs from the realm of mathematical calculus and "prose" comes into play, in Wittgenstein's view. Therefore, Wittgenstein's "non-revisionist" attitude does not apply to proofs by contradiction that rest on intended interpretations. A rigorous mathematical proof should not be affected by the problem that some intended interpretation may not refer to that to which it is intended to refer, which is a genuinely philosophical problem.

Arithmetic interpretations that paraphrase PM formulas in terms of propositions regarding natural numbers as well as meta-mathematical interpretations that interpret abbreviations of PM formulas, such as G, as propositions regarding PM formulas are interpretations of this kind. The correctness of PM is measured relative to arithmetic models specified by arithmetic interpretations. Because Wittgenstein's critique concerns the meta-mathematical interpretation of G, he questions neither the relation between the arithmetic interpretations and arithmetic models of PM expressions nor the correctness or consistency of PM. Thus, although Wittgenstein even considers the possibility that arithmetic interpretations may not refer to their intended meanings (cf. VII,§21, paragraph 5), I will not relate this sort of critique to PM formulas. As I will argue in section 4, Wittgenstein's critique of the meta-mathematical interpretation of G can be traced back to the representability of the provability of PM formulas within the language of PM.

The criticism has been put forward, e.g., by Gödel himself (cf. Wang [1987, p. 49]), that Wittgenstein compares Gödel's proof to paradoxes; cf. in particular I,§11-13. Wittgenstein seems to identify a contradiction between the provability of G and its meta-mathematical interpretation, although Gödel's syntactic proof refers to a contradiction within PM and his semantic proof can be traced back to a contradiction between the provability of G and its *arithmetic* interpretation. Thus, the provability of G seems to be shown by the proof to be incompatible with either the consistency or correctness of PM, whereas its incompatibility with the meta-mathematical interpretation of G seems to be irrelevant to the proof. Before I discuss this critique, I will consider a view on the relation between contradictions and paradoxes that Wittgenstein stated in his middle period; I take this view as a background for his discussion of contradictions in the foundations of mathematics and in meta-mathematics.

According to Wittgenstein, so-called contradictions relying on the *interpretations* of provable or contradictory formulas are paradoxes (or antinomies); cf. WVC, pp. 121f.:

> If nowadays you asked the mathematicians, '[...] Have you ever encountered a contradiction *in* mathematics?,' they would appeal to the antinomies [of set theory] in the first place, [...] Now it has to be said that these antinomies have nothing whatsoever to do with the consistency of mathematics; there is no connection here at all. For the antinomies did not arise in the calculus but in our ordinary language, precisely because we use words ambiguously. [...] Thus the antinomies vanish by means of an *analysis*, not by means of a *proof*.

This conception of the antinomies of set theory is contrary to a standard view. According to this standard view, Russell's Antinomy – "the contradiction", in Russell's words (Russell [1992]), §78 – reveals a contradiction within Frege's calculus (specifically, in Frege's axiom V) or, similarly, within a calculus of naïve set theory (specifically, in the so-called axiom schema of unrestricted comprehension, $\exists x \forall y (y \in x \leftrightarrow \varphi(y))$, abbreviated as UCAS). To understand Wittgenstein's view, one must regard $y \in x$ as an interpretation of a dyadic logical predicate rather than regarding $\in$ as a logical constant. Consequently, Russell's Paradox, as well as other paradoxes and diagonal arguments, can be conceptualized as interpretations of the following contradictory first-order formula (cf. Simmons [1993, p. 25], footnote 10):

$$\exists x \forall y (Fyx \leftrightarrow \neg Fyy) \tag{1}$$

Interpreting Fyx such that $y \in x$ yields Russell's Paradox. This is a paradox as soon as one assumes that the concept of "the set x of all sets y such that y is a member of x iff y is not a member of y", abbreviated as (R), refers to a set (namely, the set of all normal sets). Under this assumption, a set exists according to the *interpretation* of formula (1) that cannot exist according to (1), which is contradictory. According to Wittgenstein, however, an analysis of the concept (R) shows that (R) itself is inconsistent and thus does not refer to anything, including any sort of set, such as the empty set or a set of all normal sets. The calculus of first-order logic allows the negation of (1) to be proven as a theorem. Interpreting this theorem according to Russell's Paradox yields the contradiction that a formula based on a negated existential claim is provable and simultaneously makes a statement about *something*, namely, the set of all normal sets. From Wittgenstein's perspective, it is not correct to conclude that something like "the set of all normal sets" does not exist. Instead, any interpretation of a logical theorem is an empty tautology. Thus, nothing meaningful is proven. It is the assumption that interpretations of logical theorems yield meaningful propositions concerning the existence or non-existence of certain entities that must be abandoned.

The ambiguity of ordinary language in the case of Russell's Paradox is due to an ambiguous use of ordinary predicates. (R) seems to be an admissible instance of expressions of the form "the set x of objects y such that y is a member of x iff $\varphi(y)$", abbreviated as (M). Expressions of this form usually identify sets. The fact that the substitution of $\neg y \in y$ for $\varphi(y)$ yields an inconsistent concept, however, shows that no set is identified in this special case. From Wittgenstein's perspective, the problem is not inherent to the calculus but rather is a problem of intended interpretation or instantiation: not every instance of (M) yields a meaningful proposition. If one intends to identify sets by concepts, one must not conceive of *any* ordinary predicate (or propositional function) as an expression of a concept referring to a set. Instead, one should distinguish predicates expressing *material concepts* from those expressing *formal concepts*; cf. TLP 4.123, 4.126. Only the former identify sets and are admissible instances of logical predicates (and, consequently, of $\varphi(y)$ in (M) or UCAS). In the case of a formal concept such as being a tautology or contradiction, however, the question of whether some object satisfies the formal concept is a question of algorithmic decision. Because instances of (R) are inconsistent as a result of the logical form of (R), this concept is formal and, thus, does not refer to any set.

One might apply this sort of analysis not only to known paradoxes but also to proofs by contradiction, such as Cantor's proof that the set of real numbers is uncountable. This proof can also be traced back to an interpretation of (1); cf., e.g. Redecker [2006, p. 105]. However, it is controversial whether Wittgenstein indeed questions Cantor's proof as a non-mathematical "prose-proof" based on a misleading interpretation of (1) or whether he questions only certain misleading "prose-interpretations" of an unquestionable mathematical proof.[9] I cannot discuss this controversy here, nor do I wish to assume any particular understanding of Cantor's proof and Wittgenstein's critique of it in the following. I merely wish to draw an analogy between questionable proofs of contradiction and the interpretation of Cantor's proof as a "prose-proof". According to this interpretation, the contradictory assumption of an anti-diagonal sequence being a member of an enumeration of real sequences, based on which the anti-diagonal sequence itself is defined, is not even a meaningful assumption. Therefore, its negation is also not meaningful.[10] Thus, proofs of contradiction based on the interpretation of formal contradictions do not prove meaningful theorems. Consequently, they do not serve as impossibility proofs because they do not refer to any well-defined entity. The fact that all interpretations of logical theorems, i.e., all tautologies, are equivalent and thus identical in meaning makes evident that proofs of tautologies via proofs by contradiction do not prove anything specific. They do not prove different "theorems", as is intended according to their various corresponding interpretations. Wittgenstein rejects proofs by contradiction that rely on interpreting logical contradictions because they attempt to "think the unthinkable" (RFM VI, §28).

Wittgenstein's analysis of paradoxes is not restricted to interpretations of contradictions such as (1) (or provable formulas such as $\neg(1)$). One might consider, e.g., the Liar as an interpretation of the provable formula $\neg(Q \leftrightarrow \neg Q)$, abbreviated as (E). Interpreting Q as "This sentence is not true" yields a

[9] Redecker [2006] provides a detailed interpretation of the first kind; [Dawson, 2016, section IV] provides a recent and thorough analysis of the second kind. Ramharter [2001] substantiates that Wittgenstein discusses different types of contradictions, separates revisionist and non-revisionist parts of Wittgenstein's critique and explains the sense in which Wittgenstein's objection to Cantor's proof is reasonable.

[10] Negating the statement that some anti-diagonal sequence exists that is a member of the enumeration on which it is defined is not equivalent to stating the existence of an anti-diagonal sequence that is not a member of that enumeration, which is unproblematic. There is a "sober", consistent meaning of the statement that infinite sequences are not mapped to natural numbers in a one-to-one manner if one intends to refer not to an anti-diagonal sequence that is a member of the enumeration on which it is defined but rather to one that is a member of some other enumeration. In this case, the diagonal argument does not establish a proof by contradiction. Instead, it simply defines a new entity that is necessarily a member of some further enumeration of infinite sequences. From this, it follows neither that "the" set of reals is uncountable, nor, of course, that "the reals" are countable or enumerable. Given that one refers (unambiguously) to anti-diagonal sequences that are *not* members of the enumerations of sequences on which they are defined, all that one can provide are enumerations of infinite sequences with anti-diagonal sequences that are, in turn, members of other enumerations of infinite sequences with their own anti-diagonal sequences. Rather than proving that "the set" of real numbers is uncountable, it is shown that there is no reasonable "set" that encompasses the totality of all real numbers; cf. II, §33. Likewise, Wittgenstein also rejects the assertion that there exists a well-defined concept of "*all* reals" or "*the set* of reals", and consequently, he rejects the notion of a well-defined concept of "enumerating" or "counting *the* reals"; cf., e.g., II, §17, §20, §22.

contradiction between the provability of (E) and its intended interpretation, abbreviated as (L), that allows one to interpret Q as true iff Q is not true. This contradiction is not due to any inconsistency or incorrectness of the propositional calculus. Instead, it is due to the ambiguity of the intended interpretation, which interprets Q as referring to the entire sentence (L) in addition to referring to what the entire sentence (L) is about; cf. PR, p. 207f. "x is not true" does not express a material concept because in the special case of self-reference, the name in the argument position does not have an unambiguous reference.

The important point with respect to Wittgenstein's critique of Gödel's proof, however, is not the analysis of paradoxes or diagonal arguments in terms of interpretations of contradictory or provable *first-order* formulas. Instead, Wittgenstein's critique concerns the relation between interpretations and formulas of calculi in general: the formal properties of a formula (in the case of UCAS, its consistency for all concepts $\varphi(y)$ referring to sets; in the case of (1), its inconsistency; in the case of $\neg(1)$ or (E), its provability) seem to contradict a property of its interpretation (in Russell's Paradox, the property of referring to a set; in Cantor's theorem, the property of referring to a sequence; in the case of the Liar, the property of referring to a truth value). According to Wittgenstein, the assumed properties of the interpretation must be given up: one seems to refer to something, whereas, in fact, one does not.

Wittgenstein's analysis of paradoxes is a special case of how to cope with so-called "non-extensional contexts". It is not the calculus that is questioned but rather instances of formulas that do not satisfy the principle of extensionality. Wittgenstein applies this analysis to diagonal arguments, such as Russell's Paradox, Cantor's theorem and the Liar. In the "diagonal context", expressions have no reference, or at least no unambiguous reference. The provability of a formula does not imply that its instances are meaningful and true. Instead, an instance may be senseless, nonsensical or false. This does not call into question the correctness of the calculus because the correctness of a calculus is measured in terms of formal (or restricted), purely extensional semantics rather than specific instances. In particular, the correctness of a calculus, at least a calculus such as first-order logic or PM, is not measured with respect to its meta-mathematical interpretations. Wittgenstein's general point is that intended interpretations cannot serve as criteria for assessing the formal properties of a calculus because in a case of doubt, the admissibility of the intended interpretation can be repudiated. The possibility or correctness of syntactic proofs should not be judged on the basis of intended interpretations; rather, the admissibility of intended interpretations should be judged on the basis of syntactic proofs. Interpretations, in the sense of instances of formulas or their abbreviations, are related to the *application* of a calculus, not to its *justification* or its *decidability*.

This general point is independent of Wittgenstein's specific analysis of paradoxes and diagonal arguments because it is justified by well-known examples of instances of formulas that do not satisfy the principle of extensionality. As Wittgenstein states in I,§15, the general question is what criteria one regards as valid for making judgments regarding the formal properties of a calculus, such as provability. Because of his algorithmic conception of proof, Wittgenstein admits only *syntactic* criteria. His rejection of meta-mathematical unprovability proofs is based on his principal objection to deciding upon formal properties based on intended interpretations. In the following, I will demonstrate how this general point applies specifically to Gödel's semantic and syntactic undecidability proofs of G.[11]

4 Gödel's Undecidability Proof

Against the background of Wittgenstein's analysis of mathematical proofs by contradiction, it can be somewhat understood why Wittgenstein does not consider the details of Gödel's proof. He even intends "to by-pass it [Gödel's proof]" (VII, §19). Wittgenstein is not interested in the question of whether PM is undecidable. His only interest is in the question of whether a proof by contradiction that rests on the interpretation of a formal language must persuade one to give up the search for a proof of G or $\neg G$ within PM; cf. I,§14-17; VII,§22, paragraphs 3-6. Wittgenstein answers this question in the negative by comparing Gödel's undecidability proof with algorithmic unprovability proofs on the one hand and with impossibility proofs in the form of proofs by contradiction (proving, e.g., the impossibility of a set of all normal sets or an anti-diagonal sequence that is part of the enumeration based on which it is defined) on the other. According to Wittgenstein, it is obvious that the mere assumption of the provability of G cannot be reduced to absurdity on the basis of meta-mathematical interpretations. Instead, he reverses the relation: the provability of G reduces the meta-mathematical interpretations concerned, such as the interpretation of G as stating its own unprovability, to absurdity; cf. I,§8,§10,§17. According to Wittgenstein, a necessary criterion for admissible interpretations is that they must not be in conflict with the provability of the instantiated formulas (or their abbreviations). If this is the case, it becomes doubtful whether such expressions refer to that to which they seem to refer according to their intended meta-mathematical interpretations.

In part I of RFM, Wittgenstein is satisfied with this general critique, without specifically addressing the definitions Gödel adopts to represent meta-mathematical propositions within the language of PM. In part VII,§21f. of RFM, he notes that Gödel's meta-mathematical interpretation is based on an arithmetic interpretation of PM formulas. However, he also does not analyze the relation between arithmetic and meta-mathematical interpretation in detail. Nowhere

[11] The common distinction between Gödel's "semantic" and "syntactic" versions of his undecidability proof does not, of course, imply that his "syntactic proof" is a proof within PM or an algorithmic proof applying syntactic criteria. It merely means that it is based on the assumption of PM's consistency, instead of the stronger assumption of PM's correctness, in proving that the decidability of G implies the inconsistency of PM. However, this "syntactic" proof is a meta-mathematical proof involving interpretations. It should also be noted that the general relation between syntax and interpretation is a controversial issue. In this paper, I adhere to rather familiar notions that suffice to identify where Wittgenstein's critique relates to Gödel's undecidability proof.

does he identify a specific point at which Gödel's definitions imply a meta-mathematical interpretation and may fail. That is why Wittgenstein's critique has been accused of being irrelevant to the pertinent assumptions of Gödel's undecidability proof. In the following, I intend to counter this argument by relating Wittgenstein's critique to Gödel's semantic and syntactic proofs.

4.1 Semantic Proof

Gödel sketches his so-called semantic proof in section 1 of Gödel [1931] "without any claim to complete precision" (cf. p. 147). Whereas his syntactic proof presumes only the consistency of PM, his semantic proof presumes its correctness. It is often correctly noted that Wittgenstein seems to refer only to Gödel's informal introduction and thus passes over Gödel's syntactic version of the proof as well as the precise definitions used to represent provability within the language of PM. In section 4.2, I will explain how Wittgenstein's critique can be applied to Gödel's syntactic proof. First, however, let us examine how Wittgenstein's critique can be applied to Gödel's semantic proof.

To discuss Gödel's proof with respect to Wittgenstein's critique, it is crucial to distinguish the arithmetic standard interpretation of a PM formula φ, $\Im_A(\varphi)$, from its meta-mathematical interpretation, $\Im_M(\varphi)$. In contrast to $\Im_A(\varphi)$, $\Im_M(\varphi)$ is not an interpretation in terms of an instance of φ that paraphrases each logical and each arithmetic constant. Instead, $\Im_M(\varphi)$ is only an instance of an abbreviation of φ. All that is needed to acknowledge $\Im_M(\varphi)$ is that the corresponding meta-mathematical proposition must be representable (definable) within the language of PM; i.e., the following must hold in case of formulas φ with a meta-mathematical interpretation $\Im_M(\varphi)$:

$$\Im_M(\varphi) = T \text{ iff } \Im_A(\varphi) = T \tag{2}$$

In fact, Gödel's sentence G is a very long formula. Consequently, its arithmetic paraphrase is also very long. Despite this fact, its meta-mathematical interpretation "G is unprovable" or, more accurately, 'There is no y such that y is the number of a proof of the formula with the number $\ulcorner G \urcorner$' as instance of $\neg\exists y B(y, \ulcorner G \urcorner$, is very short. To be an admissible interpretation of G, this meta-mathematical interpretation $\Im_M(G)$ must satisfy the equivalence condition given in (2).

Wittgenstein does not argue for abandoning $\Im_A(G)$. Thus, in the following, let us presume the correctness of PM.

Contrary to Gödel, however, Wittgenstein does argue that $\Im_M(G)$ should be abandoned. In RFM I, he never refers to $\Im_A$ at all. In RFM VII, he distinguishes between $\Im_A(G)$ and $\Im_M(G)$ to clarify that only the Gödel figure of G[12], and not the formula G itself, appears in G; cf. VII,§21. However, like Gödel (Gödel [1931, p. 148]) and Wittgenstein himself in RFM I, Wittgenstein

[12]Wittgenstein does not speak, as is usual, of the (Gödel) *number* of a formula. Instead, he emphasizes that (Gödel) *figures* refer to formulas and that it is an open question when those figures also refer to numbers. I will argue that it is essential that Wittgenstein does not presume that figures assigned to formulas also refer to numbers in the context of the recursive

traces his understanding of Gödel's proof as a proof by contradiction to $\Im_M(G)$ in RFM VII, §21.[13] Nevertheless, just as one should not make too much out of Gödel's short version of his semantic proof (Gödel [1931, p. 149]) – cf. Rodych's quote below – one also should not reduce Wittgenstein's critique to a simplified analysis of Gödel's proof as a literal contradiction between the provability and meta-mathematical interpretation of G. The point of Wittgenstein's critique goes beyond such a narrow and trivialized understanding of Gödel's proof. Instead, it calls into question the equivalence expressed in (2) with respect to G; i.e., it challenges the following equivalence:

$$\Im_M(G) = T \text{ iff } \Im_A(G) = T \tag{3}$$

In general, Wittgenstein's critique questions whether provability is definable within PM. In the following, I will further elaborate on this point in discussing Rodych's analysis of "Wittgenstein's (big) mistake"; cf. Rodych [1999, p. 182].

According to Rodych [1999, p. 183] (as well as [Steiner, 2001, p. 263]), Gödel's formulation of his semantic proof misled Wittgenstein into assuming that $\Im_M(G)$ is essential to Gödel's proof. In the following quote, Wittgenstein's P and Gödel's $[R(q); q]$ both represent Gödel's sentence G:

> Wittgenstein [...] seems to think that a natural language interpretation of P, such as Gödel's original "the undecidable proposition $[R(q); q]$ states [...] that $[R(q); q]$ is not provable", is essential to the proof. If Wittgenstein is relying exclusively upon Gödel's original paper, the fault is not entirely his own, for Gödel himself says that "[f]rom the remark that $[R(q); q]$ says about itself that it is not provable it follows at once that $[R(q); q]$ is true". The critical point, however, is that *all of this is irrelevant* – this is Wittgenstein's mistake. To show, *meta-mathematically*, that P is true if it is unprovable, we need only show that a particular number-theoretic proposition, say $[R(q); q]$, is true *iff* a particular number-theoretic proposition, say $[R(q); q]$, is unprovable (in Russell's system). It is entirely unnecessary to give $[R(q); q]$ a *natural language* interpretation to establish the bi-conditional relationship.

Prima facie, Rodych is correct. All that is relevant for Gödel's proof is that the following is true given the correctness of PM:

$$\Im_A(G) = T \text{ iff } \nvdash G \tag{4}$$

However, Gödel must prove this equivalence. The question is whether this is possible without presuming (3). In the following, I will show that this is not

definition of "x is a proof of y", and this is why I use the phrase "Gödel figures" instead of the usual term "Gödel numbers".

[13]Cf. VII, §22: "If we had then derived *the arithmetical proposition* from the axioms according to our rules of inference, then *by this means* we should have demonstrated its derivability, but we should also have proved a proposition which, by that translation rule, can be expressed: this arithmetical proposition (namely ours) is not derivable. [...] If we now read the constructed proposition (or the figures) as a proposition of mathematical language (in English, say) then it says the opposite of what we regard as proved."

the case. According to Gödel's own presentation as well as appropriate reconstructions thereof, Gödel's proof presumes the definability of provability and thus presumes (3). Gödel mentions explicitly that his semantic and syntactic proofs both presume the definability of meta-mathematical concepts such as "formula", "proof array", and "provable formula"; cf. Gödel [1931, pp. 147, 151; note also footnote 9]. An essential part of his complete proof contains precisely these definitions. The semantic and syntactic proofs differ only in that in the syntactic proof, the assumption of correctness is replaced with the "much weaker" inconsistency assumption (Gödel [1931, p. 151]).

Although Gödel clearly distinguishes between recursively defined functions (and relations) and their representation within the language of PM, he does not express this distinction in his notation; cf. Gödel [1931, pp. 156-158]. However, to see where Wittgenstein's critique becomes relevant, it is expedient to distinguish recursively defined functions/relations from their corresponding PM expressions, as is conventional in modern reconstructions of Gödel's proof.

To make implicit assumptions concerning the interpretation of PM formulas explicit within Gödel's proof, one must conceptualize recursively defined functions in a syntactic manner. Thus, recursive functions relate *figures* rather than numbers, cf. Shapiro [2017, pp. 1f.]. A computer operates on a purely syntactic level; it relates nothing but binary figures (as input and output). Any interpretation of this relation is human in origin. The usual number-theoretic (arithmetic) interpretation of a recursive function, strictly speaking, already extends beyond a purely mechanical execution of recursive definitions. To relate Wittgenstein's critique to Gödel's proof, it is crucial to differentiate between a mechanical, purely syntactic manipulation of figures and the interpretation of those figures. The mere use of figures does not imply a specific interpretation; cf. VII,§22:

> But it must of course be said that that sign [i.e., the Gödel figure for G] need not be regarded either as a propositional sign or as a number sign. – Ask yourself: what makes it into the one, and what into the other?

One might think about, e.g., the recursive definition of truth functions. Here, '0' and '1' refer to truth values. Within this context, one operates with figures as one does with propositions, not as one does with numbers. The null function, for example, allows any figure to be identified with '0' regardless of whether this is justified by a particular arithmetic operation with numbers, such as multiplication by 0. Recursively defined functions do more than what one might call "computing numbers". This property is what makes them so useful because it enlarges the realm of what is computable by purely mechanical means. For example, logical relations between formulas might be computable in this manner. From this, however, one should not be led to mistakenly conclude that operating on figures in accordance with recursive definitions is identical to operating on numbers in arithmetic. The computation of figures is one thing; the interpretation of such computation is another.

Hence, although it is common to interpret recursively defined functions arithmetically as number-theoretic functions, one should not presume this interpretation in the case of Gödel's definition for proof if one intends to explain how Wittgenstein's critique relates to Gödel's proof. In Gödel's recursive definition for proof, (Gödel) figures refer to initially given formulas or sequences of formulas. Let $B^*(m,n)$ be the recursive relation whose characteristic function decides, for a Gödel figure m of a sequence of formulas and a Gödel figure n of a formula, whether m is the figure of a proof of the formula with figure n. Then, the recursive relation $B^*(m,n)$ can be interpreted meta-mathematically through the sentence "The sequence of formulas with Gödel figure m is a proof of the formula with Gödel figure n". Let us abbreviate this as $\Im_M(B^*)$. Such a recursive definition of the relation of proof is an essential part of Gödel's proof. For the proposed understanding of Wittgenstein's critique, it should not be questioned but should rather be presumed.

Regarding the question of the definability of "x is a proof of y" within the language of PM, one must prove that the meta-mathematical interpretation of the recursive relation $B^*(m,n)$ is representable within PM by a predicate $B(m,n)$. That is, the following must hold for all m and all n:

$$\Im_M(B^*(m,n)) = T \text{ iff } \Im_A(B(m,n)) = T \tag{5}$$

In fact, it is proven by Gödel and, in greater detail, in modern reconstructions of his proof that any recursive function is representable (definable) within the language of PM (or PA). However, this proof refers to an arithmetic (not a meta-mathematical) interpretation of recursive functions; cf. Gödel [1931, pp. 182-184, proof of Theorem VII], and, e.g., Smith [2007, pp. 109-114], proof of Theorem 13.4, and Boolos [2003, pp. 199-204], chapter 16.1. The proof consists of a definition of an effective formalization procedure, formalizing any n-adic recursive relation $R^*(x_1,\dots x_n)$ in terms of an n-adic predicate $\phi(x_1,\dots x_n)$ within the language of PM (or PA) such that the following holds:

(D): For all $x_1,\dots x_n$:
$\Im_A(R^*(x_1,\dots x_n)) = T$ iff $\Im_A(\phi(x_1,\dots x_n)) = T$.

Consequently, with regard to B^*, the following is proven for all m and n:

$$\Im_A(B^*(m,n)) = T \text{ iff } \Im_A(B(m,n)) = T \tag{6}$$

To prove (5) from (6), $\Im_A(B^*(m,n))$ must be correlated with $\Im_M(B^*(m,n))$. Thus, it must be presumed that any arbitrary Gödel figures m and n refer not only to formulas or sequences of formulas but also to numbers in $B^*(m,n)$ (= presumption PS). Then, with the presumption of (5), and thus (PS), the equivalence (3) and(4) become provable. If (5) were not valid, (3) would also not be valid and there would be no reason why $\Im_A(G) = T$ should be correlated with $\nvdash G$ as (4) states. Consequently, Wittgenstein's plea for abandoning $\Im_M(G)$ given $\vdash G$ implicitly questions presumption (PS).

Thus, Wittgenstein's questioning of $\Im_M(G)$ (or, more precisely, of the equivalence (3)) can be traced back to an open question within Gödel's proof, namely, how to infer $\Im_A(B^*) = T$ from $\Im_M(B^*) = T$. Gödel does not ask, as Wittgenstein instructs the reader to do in VII, §22 (see above), what makes a Gödel figure a sign for a formula (or a sequence of formulas) and what makes it a sign for a number. The possibility of admissible interpretations depends on the use of the corresponding signs in the corresponding contexts. Just as one might ask whether P in "P is false" refers to the propositional sign "P is false" or to the object that the statement "P is false" is about, the reference of the Gödel figure of G in $\Im_A(B^*)$ and, consequently, the interpretation of G remain ambiguous.

This critique is directly relevant to Gödel's proof. At the same time, it extends Wittgenstein's analysis of paradoxes and proofs of contradiction to Gödel's proof. According to Wittgenstein, paradoxes are based on ambiguity within assumptions of interpretation. In Gödel's proof, this ambiguity lies in the interpretation of Gödel figures as figures of PM expressions and of numbers. The fact that meta-mathematical and arithmetic interpretations are equivalent in many (most) cases does not mean that they are so in *all* cases. As paradoxes show, ambiguity of reference may induce contradictions within diagonal contexts. G is constructed such that the Gödel figure of G, on the one hand, refers to the entire formula, according to its meta-mathematical interpretation $\Im_M(G)$, and, on the other hand, simultaneously refers to a number as an object of predication within G, according to the arithmetic interpretation $\Im_A(G)$. The correctness of PM implies only that the provability of G is correlated with $\Im_A(G)$; it does not imply that it is correlated with $\Im_M(G)$. Thus, it is possible that $\vdash G$, $\Im_A(G) = T$ and $\Im_M(G) = F$ may be correlated. Wittgenstein argues that one could, indeed, conclude $\Im_M(G) = F$ from the assumption $\vdash G$, thereby presuming the correctness of PM and, thus, $\Im_A(G) = T$. Instead of inferring the incorrectness of PM from the hypothetical assumption of $\vdash G$, he suggests to infer that in the PM formula G, the Gödel figure $\ulcorner G \urcorner$ of G refers not to the PM formula G but rather to a number, whereas in the corresponding recursive relation $B^*(k, \ulcorner G \urcorner)$ (where k is the Gödel figure of the proof of G), $\ulcorner G \urcorner$ refers to the PM formula G and not to a number.

4.2 Syntactic Proof

To replace the assumption of the correctness of PM with the *weaker* assumption of the consistency of PM, Gödel proves not only the definability of recursive functions within PM, i.e., (D), but also the *stronger* theorem that recursive functions are also *captured* within PM (or PA). That is, he proves the following:[14]

(S) For all $x_1, \dots x_n$:
if $\Im_A(R^*(x_1, \dots x_n)) = T$, then $\vdash \phi(x_1, \dots x_n))$, and
if $\Im_A(R^*(x_1, \dots x_n)) = F$, then $\vdash \neg\phi(x_1, \dots x_n))$.

[14]Cf. Gödel [1931, p. 170], Theorem V; Smith [2007, p. 116], Theorem 13.6; and Boolos [2003, p. 212], Theorem 16.6(a))). In fact, it is sufficient to presume the axioms of Robinson's arithmetic Q and thus refer to $\vdash_Q$ instead of $\vdash_{PM}$ (i.e., $\vdash$; cf. footnote 4).

Wittgenstein's critique thus also relates to Gödel's syntactic version of his proof because the proof of (S) is based on the proof of (D).[15]

Gödel [1931, p. 177], Steiner [2001, p. 259] and Rodych [1999, p. 182] emphasize that Gödel's syntactic proof is "constructive". This claim is essentially grounded on the fact that recursive relations are translatable into PM predicates by means of a mechanical procedure. This fact also seems to be the basis for Gödel's reaction to Wittgenstein, in which he insists that his undecidability theorem has nothing to do with any paradox but rather is a "mathematical theorem within an absolutely uncontroversial part of mathematics (finitary number theory or combinatorics)" (Wang [1987, p. 49]).[16] For the same reason, Rodych and Steiner seem to infer that Gödel's syntactic proof does not rely on assumptions regarding meta-mathematical interpretation.[17] However, this is not correct. The proof that a procedure for translating recursive relations is *correct* presumes that $\Im_M(B^*)$ is equivalent to $\Im_A(B^*)$ in all cases. If this is not the case, it may well be that proving G, and thus generating $B^*(k, \ulcorner G \urcorner)$ and, consequently, $B(k, \ulcorner G \urcorner)$, does not imply that $\Im_A(B(k, \ulcorner G \urcorner)) = T$. Thus, when PM's correctness is presumed, it does not imply that $B(k, \ulcorner G \urcorner)$ and its existential generalization are provable. (D) and (S) simply do not apply if (5) and, thus, (3) do not hold. If one assumes (PS) without reservation, one misses a possible source of error that also underlies paradoxes. Effective translations cannot exclude potential ambiguities that threaten to give rise to inconsistent interpretations. Signs do not have fixed meanings independent of their use in specific contexts. Figures are no exception. Shifts of context may induce shifts in reference. If one does not consider this, ambiguities may arise that lead to the questionable assumption that the concept of proof is representable (definable) or even captured within the language of PM.

According to this interpretation, Wittgenstein notes that Gödel's undecidability theorem presumes not only PM's correctness or consistency but also presumption (PS). In particular, it is based on the assumption that in the case of G's provability, the figure $\ulcorner G \urcorner$ refers not only to a formula but also to a number

[15] Cf. the proofs of the theorems mentioned in footnote 14.

[16] However, it should be noted that Gödel himself compares his proof to epistemological paradoxes in Gödel [1931, p. 149, and footnote 14]. Furthermore, Gödel is well aware that his undecidability theorem is based on the representation of meta-mathematical concepts within the language of PM. His reaction to Wittgenstein makes evident that Gödel trivializes the point that is called into question from Wittgenstein's perspective. Whereas presumption (PS) seems trivial to Gödel, it induces potential ambiguities according to Wittgenstein.

[17] Rodych [1999, p. 182]:

> [. . .] it is just a number-theoretic 'fact' that an actual proof of 'P' would enable us to *calculate* the relevant Gödel numbers and thereby arrive at '∼ P' by existential generalization.

Steiner [2001, p. 259]:

> Gödel's theorem [. . .] is both finitistic and constructive. That is, Gödel gives what amounts to a remarkable computer program which transforms any 'Russellian' proof of 'the Gödel sentence' into a proof of its negation[. . .].

Cf. also Steiner [2001, p. 262]. In fact, computer programs that define B^* and translate $B^*(m, n)$ into $B(m, n)$ already exist; cf. Shankar [1997], Connor [2005], Harrison [2009] and http://tachyos.org/godel.html. Such programs also generate G.

within $B^*(k, \ulcorner G \urcorner)$. Gödel might judge any number-theoretic interpretation of any recursively defined function to be "absolutely uncontroversial". According to Wittgenstein, however, only algorithmic proofs are uncontroversial. Consequently, he is not forced to give up the search for algorithmic proofs on the basis of a proof that involves assumptions concerning the interpretation of formal expressions (including figures). This is particularly true if intricate interpretations are involved, such as $\Im_M(B^*(k, \ulcorner G \urcorner))$ and $\Im_A(B^*(k, \ulcorner G \urcorner))$ and, consequently, $\Im_M(G)$ and $\Im_A(G)$, for which arithmetic and meta-mathematical interpretations must be correlated in the case of diagonalization. In this case, any meta-mathematical interpretation should be restricted to a language different from PM (or PA), e.g., the language of recursive functions, because there is no guarantee that arithmetic and meta-mathematical interpretations are isomorphic in the diagonal case; it may well be that only the intended meta-interpretation of $B^*(k, \ulcorner G \urcorner)$ and the intended arithmetic interpretation of $B(k, \ulcorner G \urcorner)$ refer to that to which they are intended to refer.

One could assess the definability of proof and provability within the language of PM by asking whether, for all m and n, $\Im_M(B(m,n)) = T$ is indeed correlated with a respective proof. Gödel, however, does not measure definability by provability but rather measures provability on the presumption of definability. Wittgenstein inverts this relation. According to my attempt to make "Wittgenstein's objection" reasonable, he advocates for interpreting Gödel's proof not as an undecidability proof but as an indefinability proof. In contrast to Gödel with regard to undecidability, Wittgenstein does not claim that the indefinability of provability is proven. All he claims, according to my interpretation, is that the undecidability of G is not proven as long as Gödel's proof can also be interpreted as a proof of indefinability. Thus, Wittgenstein's critique by no means refutes Gödel's undecidability proof of G. His critique rather challenges the conviction that Gödel's proof is a compelling reason for giving up the search for decision procedures.

Abbreviations for Wittgenstein's Works

LFM: *Lectures on the Foundations of Mathematics*, The University of Chicago Press: Chicago 1989.

RFM: *Remarks on the Foundations of Mathematics*, M.I.T.: London 1967.

PG: *Philosophical Grammar*, University of California Press: Berkeley 1978.

PR: *Philosophische Remarks*, The University of Chicago Press: Chicago 1975.

TLP: *Tractatus logico-philosophicus*, Routledge: London 1974.

WVC: *Wittgenstein and the Vienna Circle*, Basil Blackwell: Oxford 1979.

References

Anderson, A.R.: "Mathematics and the 'Language Game"', *Rewiew of Metaphysics* 11, 1958, 446-458.

Bays, T.: "Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc.", online draft: http://www3.nd.edu/ tbays/papers/wnp2.pdf.

Berto, F.: "The Gödel Paradox and Wittgenstein's Reasons". *Philosophia Mathematica 3 (17)*, 2009, 208-219.

Berto, F.: *There's Something About Gödel: The Complete Guide to The Incompleteness Theorem*, 2009, Wiley Blackwell, New Jersey.

Bernays, P., "Comments on Ludwig Wittgenstein's Remarks on the Foundations of Mathematics", *Ratio* 2.1, 1959, 1-22.

Boolos, G.S., Burgess, J.P., Jeffrey, R.C.: *Computability and Logic*, 4th edition, Cambridge University Press, Cambridge, 2003.

O'Connor, R.: "Essential Incompleteness of Arithmetic Verified by Coq", *Lecture Notes in Computer Science* 3603, 245-260.

Dawson, R.: "Wittgenstein on Set Theory and the Enormously Big", *Philosophical Investigations* 39.4, 313-334.

Dummett, M.: "Wittgenstein's Philosophy of Mathematics", *Philosophical Review* 68, 1959, 324-348.

Floyd, J.: "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: Dordrecht, 1995, 373-425.

Floyd, J. & Putnam, H., "A Note on Wittgenstein's 'Notorious Paragraph' about the Gödel Theorem", *The Journal of Philosophy XCVII*, 11, 2000, 624-632.

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

Floyd, J. &: Putnam, H.: " Bays, Steiner, and Wittgenstein's 'notorious' paragraph about the Gödel Theorem", *The Journal of Philosophy* 103.2, 2006, 101-110.

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

Goodstein, R.L., "Critical Notice of *Remarks on the Foundations of Mathematics*", *Mind* LXVI, 549-553.

Harrison, J.: *Handbook of Practical Logic and Automated Reasoning*, Cambridge University Press, Cambrdige, 2009.

Kreisel, G., "Wittgenstein's *Remarks on the Foundations of Mathematics*", British Journal for the Philosophy of Science 9, 1958, 135-57.

Martin, G.E. (1998): *Geometric Construcktions*, Springer: New York.

Matthíasson, Á.B.: *A Chalet on Mount Everest: Interpretations of Wittgenstein's Remarks on Gödel*, online draft: https://www.illc.uva.nl/Research/Publications/Reports/MoL-2013-26.text.pdf

Mühlhölzer, F.: "Wittgenstein and the regular Heptagon", *Grazer Philosophische Studien 62*, 215-247.

Priest, G.: "Wittgenstein's Remarks on Gödel's Theorem", in: Kölbel, M. und Weiss, B.: *Wittgenstein's Lasting Significance*, Routledge, London, 206-225.

Ramharter, E.: "Are all contradictions equal? Wittgenstein on confusion in mathematics", in: Löwe, B. and Müller, T.: *Philosophy of Mathematics: Sociological Aspects and Mathematical Practice*, College Publications, London, 293-306.

Redecker, C.: *Wittgensteins Philosophie der Mathematik*, Ontos, Frankfurt, 1996.

Rodych, V.: "Wittgenstein's Inversion of Gödel's Theorem", *Erkenntnis* 51, 1999, 173-206.

Rodych, V.: "Wittgenstein on Gödel: The Newly Published Remarks", *Erkenntnis 56*, 2002, 379-397.

Russell, B.: *The Principles of Mathematics*, Routledge, London, second edition, 1992.

Shankar, N.: *Metamathematics, Machines, and Gödel's proof*, Cambridge University Press, Cambridge, 1997.

Shanker, S.G.: "Wittgenstein's Remarks on the Significance of Gödel's Theorem". in S. G. Shanker (Ed.), *Gödel's Theorem in Focus*, Croom Helm, New York, 1988, 155-255.

Shapiro, S.: "Computing with Numbers and Other Non-syntactic Things: *De re* Knowledge of Abstract Objects", *Philosophia Mathematica*, online first: https://doi.org/10.1093/philmat/nkx009, 1-14.

Simmons, K.: *Universality and the Liar: an Essay on Truth and the Diagonal Argument*, Cambridge University Press, New York, 1993.

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

Smullyan, R.M: *Gödel's Incompleteness Theorems*, Oxford University Press, Oxford, 1992.

Steiner, M.: "Wittgenstein as His Own Worst Enemy: The Case of Gödel's Theorem", *Philosophia Mathematica* 9, 2001, 257-279.

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