A Logical Refutation of Wittgenstein's Early Philosophy of Logic

XXX, YYY

Abstract

An essential feature of Wittgenstein's early philosophy of logic is the conjecture of a positive solution to the problem of whether logical truth in first-order logic is decidable (*Entscheidungsproblem*). It is often argued that the proofs of a negative solution to the *Entscheidungsproblem* presented in the 1930s refuted Wittgenstein's early conception of logic. However, Wittgenstein's philosophy of logic does not share the principles presumed by the proof strategies of those undecidability proofs. Therefore, we provide a purely logical refutation of Wittgenstein's early philosophy of logic that does not make use of assumptions that Wittgenstein does not share. Wittgenstein asserted that logical properties of first-order logic are decidable on the basis of patterns in a suitable notation for first-order logic. We explain why this conviction fails.

1 Introduction

An essential and provoking feature of Wittgenstein's early philosophy of logic (WPL) is his conviction that first-order logic (FOL) is decidable. Since it is a common theorem of mathematical logic that FOL is undecidable[1], WPL is judged to be refuted (cf., e.g., Anscombe [1959], p. 137; Black [1964], p. 319; Fogelin [1976], p. 75; Sundholm [1990], p. 60; editor's comment in CL, p. 52; Landini [2007], pp. 118–120; Potter [2009], pp. 181–183). This paper examines this judgement and argues that additional effort must be made to refute Wittgenstein's conviction because Wittgenstein does not accept presumptions of a metalogical undecidability proof. We, thus, provide an undecidability proof solely based on assumptions that do not go beyond automated proof search and, thus, intend to explain to a Wittgensteinian why the view of WPL is not applicable to the whole realm of FOL.

Since we are concerned with the *Entscheidungsproblem*, we abstain from discussing special features of Wittgenstein's early logic. Specifically, we do not

[1] This theorem is called the "Church–Turing theorem" because Church [1936] and Turing [1936/37] each proved it independently in 1936. In doing so, they used different, although equivalent, notions of computability. Theorems IX and X of Gödel [1931] were already very near to Church's proof, but only Turing's later analysis of computability seemed to convince Gödel; cf. Kripke [2013].

consider questions about the application of logic or of logical analysis of language. More importantly, we do not consider the peculiarities of Wittgenstein's early logic; on this topic, cf. Rogers/Wehmeier [2012], Weiss [2017], and Lampert/Säbel [2021]. In particular, these concern Wittgenstein's N-operator, his interpretation of identity and the exclusive interpretation of quantifiers. We do not intend to contribute to exegetical controversies concerning Wittgenstein's early conception of logic. Instead, we take it for granted that Wittgenstein thought traditional FOL to be decidable, independent of philosophical controversies about its semantics and alternatives he might have envisaged. Since we do not consider identity, we refer to FOL without identity. As is well known, the *Entscheidungsproblem* for FOL with identity can be reduced to the *Entscheidungsproblem* for FOL without identity.

We first substantiate Wittgenstein's conviction of a positive solution to the *Entscheidungsproblem* (section 2) before we discuss the question of the extent to which this conviction is refuted by the Church–Turing theorem (section 3). The main part of our paper (section 4) then expounds a purely logical refutation of WPL that is based on assumptions Wittgenstein shares. The last section, section 5, concludes that WPL is besotted by a vision of a decision procedure in terms of pattern detection that, however, is realizable only for fragments of FOL.

2 The *Entscheidungsproblem* and WPL

Following Russell, Wittgenstein calls propositions that are true due to their logical vocabulary "logical propositions". They are equivalent to logical theorems or logically valid formulas. TLP 6.113 mentions the key idea of Wittgenstein's philosophy of logic:

> It is the peculiar [characteristic] mark of logical propositions that one can recognize that they are true from the symbol alone, and this fact contains in itself the whole philosophy of logic.

TLP 6.113 echoes Wittgenstein's explanation of this idea in a letter to Russell from 1913 (CL, letter 32):

> It is the peculiar [characteristic] (and *most* important) mark of *non*-logical propositions that one is *not* able to recognize their truth from the propositional sign alone. If I say, for example, 'Meier is stupid', you cannot tell by looking at this proposition whether it is true or false. But the propositions of logic – and only they – have the property that their truth or falsity, as the case may be, finds its expression in the very sign for the proposition. I have not yet succeeded in finding a notation for identity that satisfies this condition; but *I have* NO *doubt* that it MUST [upper case added] be possible to find such a notation.

This quote makes clear that Wittgenstein did not assume that one could somehow identify logical propositions by means of some ordinary notation for FOL. Instead, to identify logical propositions, ordinary notation must be converted into a suitable logical notation that alone provides syntactic criteria for logically true propositions. Converting a formula into a suitable notation is a finite, mechanical process. We abbreviate the idea of a decision procedure for FOL formulas consisting of conversion into a suitable notation that provides syntactic criteria for logical properties as DPW, where "DP" is short for "decision procedure" and "W" for "Wittgenstein".

Wittgenstein goes on to explain his *ab*-notation as a notation that provides a general criterion for logical propositions in the case of propositional logic, and he purports that "it is easy to see that it MUST [upper case added] also apply to all others" (CL, letter 30), whereby he explicitly refers to logical propositions of *Principia Mathematica* and, thus, to traditional FOL. In NL, p. 95ff., he illustrates the application of his *ab*-notation for most simple first-order formulas.

We will not go into the details of Wittgenstein's *ab*-notation. Lampert [2017a] showed how this notation can be applied to the fragment of FOL without dyadic connectives in the scope of quantifiers.[2] Wittgenstein, however, was not interested in specifying the details of a notation satisfying DPW. It sufficed for him to envisage its realization for fragments of FOL by means of his *ab*-notation and to stipulate that it or some alternative notation (cf. CL, letter 32, p. 52) must be realizable for the whole realm of FOL. He was concerned with solving the philosophical problem of distinguishing logical truth from other truths by DPW. His reasoning is best understood as the postulation of a position that he considered necessary to solve this problem. For him, a suitable notation for deciding FOL was a normative *postulate* that alone could solve the problem of how to explain logical truth (TLP 6.1222(2), 6.1223):

> Not only MUST [upper case added] a proposition of logic be irrefutable by any possible experience, it MUST [upper case added] also be unconfirmable by any possible experience.
>
> Now it becomes clear why people have often felt as if it were for us to '*postulate*' the 'truths of logic'. The reason is that we can postulate them in so far as we can postulate an adequate notation.

WPL stands in contrast to the distinction between syntax and semantics that is taken as standard in modern mathematical logic. According to mathematical logic, semantics assigns extensions to the nonlogical vocabulary, and logical truth is defined as truth in all interpretations (logical validity). Any FOL calculus must meet this standard for being correct and complete. Thus, syntactic transformations of signs are not seen as something that is able to define logical truth; they are merely something that must be adjusted to a prior semantic definition. Since extensions come first in this view, FOL becomes dependent

[2] Landini [2021] also considers the application of the *ab*-notation to FOL. However, he does not spell out how to realize Wittgenstein's idea to reduce all equivalent formulas to one and only one representative in the *ab*-notation for a decidable fragment of FOL.

on set theory. This conception of logic is powerful. It serves as the foundation of modern mathematics and leads to far-reaching metamathematical theorems (including the Church–Turing theorem).

However, it characterizes logical truth by a universal statement ("truth in *all* interpretations") and thus does not distinguish it *by form* from other universal statements that are not logical truths. Logical truth seems to be a matter of extensions, as any arbitrary truth of sentences. From a philosophical point of view that insists on conceptual analysis and intends to define logical truth by some kind of necessity, this is unsatisfactory. Wittgenstein explicitly intends to distinguish logical truth, as "essential" or "necessary truth", from "accidental" truth (TLP 6.1232, 4.464, 6.375) and rejects defining truth by universal validity (TLP 6.1231); cf. Etchemendy [1990] for a more recent prominent advocate of this critique. A common and stringent reaction to the critique of defining logical truth by truth in all interpretations is to abandon the philosophical ambition of a conceptual analysis that attempts to do justice to a dubious modal claim of logical truth and to instead be satisfied with an adequate extensional model for logical truth based on the assignment of extensions; cf., e.g., Ray [1996] and Shapiro [1998] in reaction to Etchemendy. Our argument in this paper intends to support this reaction by demonstrating where and why logical truth in FOL becomes an indispensable matter of extensions.

WPL, however, intends to solve the notorious problem of specifying the necessity of logical reasoning by defining logical truth in terms of nothing but syntactic transformations of signs. Syntax comes first and specifies conditions for the truth and falsehood of complex propositions or, as we will say in order to relate Wittgenstein's position to the standard view, syntax defines possible models and countermodels (extensions, in short). To do so, Wittgenstein must postulate that it is indeed possible to transform initial formulas into formulas of a proper notation such that the form of the resulting formulas is shared with any possible extension, thus making it possible to distinguish between models (true interpretations) and countermodels (false interpretations) by means of properties of their symbolization. Syntax, therefore, is prior to semantics, and logical properties such as logical truth are identified through syntactic criteria. Logical truth becomes "truth by virtue of form" in the sense that all only formulas of an equivalent class are reduced to one an the same expression in an ideal notation.

Judged from today, this may seem an idiosyncratic view of a philosopher's mind. However, it becomes a fascinating and reasonable attempt to pursue a philosophical project as soon as one comes to see how it can be realized for fragments of FOL. We will exemplify this for several fragments in section 4. For now, the reader may consider truth tables as an example of a decision procedure that distinguishes between models and countermodels by notation. From Wittgenstein's point of view, a truth table must be seen as a proper notation that provides syntactic criteria for deciding upon the logical properties of an initial formula that this formula itself does not provide. For example, logical truth is identified by the fact that no 'F' occurs below the main connective.

To solve a problem, however, it is insufficient to merely postulate a method of solving it. One must also show that this method does not solve only paradigmatic cases but solves the problem entirely. Therefore, one might admit that solving the *Entscheidungsproblem* would solve the problem of how to define logical truth in FOL and still insist on a proof that DPW can be realized for the whole realm of FOL. Contrary to Wittgenstein, we thus do not accept the mere postulation of DPW, even if it would turn out to be true that this would be required to solve all-important problems in the philosophy of logic. Instead, we wish to evaluate the feasibility of DPW and, thus, the philosophical project of defining logical truth without referring to logical validity.

3 Is DPW Refuted by Metalogic?

Like many others before, Landini [2007], p. 118, simply states that DPW is refuted by the negative solution to the *Entscheidungsproblem*:

> The undecidability of quantification logic is a significant blow to Wittgenstein's conception of logic ... it undermines Wittgenstein's hope of finding a notation in which all and only logical equivalents have one and the same representation.

The proof of the Church–Turing theorem is a masterpiece of metalogic. Wittgenstein, however, was sceptical about metamathematics (including metalogic). This scepticism was already evident in WPL. Wittgenstein rejected measuring formal systems in terms of semantics. According to his view, logic is autonomous (TLP 5.473). He laid down as a "fundamental principle [...] that whenever a question can be decided by logic at all it MUST [upper case added] be possible to decide without more ado"; considering anything beyond logic to solve a problem of logic "shows that we are on a completely wrong track" (TLP 5.551). Connected with the principle that logic is autonomous is Wittgenstein's view that we cannot go wrong in logic because anything that is possible in logic is also legitimate, or permitted; cf. NB 22.8.14. As a consequence of this view, he thought that paradoxes that arise from self-reference (such as Russell's Paradox or the Liar Paradox) do not need to be solved by *forbidding* them to be expressed (as Russell does with his Theory of Types) but rather can be solved by showing that it is *impossible* to express them (TLP 3.33-3.333). This Wittgenstein intends to make manifest by a suitable logical notation for FOL, since it identifies conditions of truth by its syntax and thus ensures that whatever is expressible in FOL has well-defined truth conditions.

To understand Wittgenstein's scepticism towards metalogical proofs, one must recognize that diagonal functions that permit self-reference occur not only in paradoxes but also in metalogical proofs. Let us first consider paradoxes before explaining why Wittgenstein's conviction casts doubt on the proof of the Church–Turing theorem.

Consider the interpretation of the formula P by "P is not true" or the interpretation of Fx by "x is not true". Such an intended, self-referential interpretation is not admissible (or, in Wittgenstein's words, is not possible).[3] The usual

[3] Cf. PR §171[8], where Wittgenstein refers by "the previous philosophy of logic" to Frege's

argument against such intended interpretations is that they violate principles of semantics laid down for FOL. The former interpretation above does not assign a truth value to P, and the latter does not assign a well-defined set to Fx. If semantic principles are violated, applying the inference rules of FOL yields fallacies; FOL's syntax applies only to propositions obeying FOL's semantics. From this argument, however, the general question arises of when interpretations are admissible. Interpretations by diagonal functions are a special and relevant case in which this question arises.

A sceptical position that does not accept semantics as standard poses the following question: How can one know that certain intended interpretations do not obey the principles of FOL's semantics and, thus, are not admissible? This question becomes relevant as soon as hypothetical reasoning comes into play. In this case, knowledge of truth values cannot be assumed. Moreover, the question arises of whether the unknown truth values of certain propositions are well defined. This question becomes relevant in metalogical proofs based on intended interpretations that make use of diagonal functions. In this case, a sceptic may turn the question around and question the admissibility of the intended interpretations involved rather than accepting the consequences of the proof. If the provability of a formula and the truth value of its intended interpretation clash, this may be taken as a criterion for asserting the inadmissibility of the intended interpretation rather than for reducing some hypothetical assumption to absurdity.

There is evidence that Wittgenstein's later critique of metamathematics is based on this sceptical attitude; cf. Anonym3 [000] for more details. However, we neither intend to enter this discussion here nor wish to evaluate this sceptical attitude. For our argument, it suffices to acknowledge that one can raise the doubt that Church's or Turing's proof is based on assumptions that DPW is not ready to accept. Let us illustrate this in the case of Turing's proof.

Turing's proof can be summarized as follows (cf. Turing [1936/37], section 11).

Turing's Thesis: Any mechanical procedure is reducible to a Turing machine (TM).

Halting Theorem: The halting problem is unsolvable (proof by means of the diagonal argument).

Turing's Lemma: A TM, started with I, halts iff $Un(TM, I)$ is provable ($\rightarrow$: syntactic proof by means of proof schema; $\leftarrow$: semantic proof by means

and Russell's mathematical logic in connection with the admission of the possibility of self-reference:

> The Cretan liar paradox could also be set up with someone writing the proposition: 'This proposition is false'. The demonstrative takes over the role of 'I' in 'I'm lying'. The basic mistake consists, as in the previous philosophy of logic, in assuming that a word can make a sort of allusion to its object (point at it from a distance) without necessarily going proxy for it.

Cf. also WVC, p. 122, where Wittgenstein states that antinomies in the foundations of mathematics arise through ambiguities in the interpretation, not in the syntax.

of the correctness of FOL, i.e., any (admissible) interpretation of a provable formula is true + application to the intended interpretations $\Im_i$ of $Un(TM, I)$).

Turing's Theorem: FOL is undecidable (proof by reductio).

Let us abbreviate the assumed machine that decides FOL as "FOL". The machine FOL returns 1 in the case of provability and 2 otherwise. To illustrate the diagonal argument as expressed within the language of FOL, one can make use of a dithering machine D that dithers if started with 1 and halts if started with 2; cf. Boolos et al. [2003], p. 39. We abbreviate the combination of the hypothetical machine FOL and the known machine D as $FOLD$. According to Turing's proof, one can reduce the existence of FOL to absurdity by the following argument:

Turing: FOL cannot be decided because if it were decidable, it would be possible to express the TM $FOLD$ applied to its own number $I(DFOL)$ by a formula $UN(FOLD, I(FOLD))$; then, however FOL decides $UN(FOLD, I(FOLD))$, this would contradict *Turing's lemma*.

Since Wittgenstein is suspicious of hypothetical reasoning based on a diagonal argument that makes use of intended, self-referential interpretations, he can reply as follows.

Wittgenstein: This shows not that FOL is undecidable but rather that the intended self-referential interpretation $\Im_i(UN(FOLD, I(FOLD)))$ is not admissible.

According to Wittgenstein, semantics is the source of philosophical problems of logic. Philosophical problems of logic are to be solved by referring to nothing but syntactic manipulations of signs. According to this view, metalogical proofs proving the *absurdity* of the *hypothetical* assumption of a decision procedure for FOL on the basis of *semantic* notions (intended interpretations, correctness) and a *diagonal* argument cannot serve as the standard for the limits of a purely mechanical and syntactic approach. Instead, the syntax of the ideal notation is the standard for what counts as admissible interpretations.

Of course, this reaction is not an argument against Turing's proof. Rather, it articulates a position that does not accept the principles of Turing's proof strategy. One might say that a philosopher should not question the strategies of established proofs in science in order to save philosophically motivated postulates. However, this may convince a scientist, but not a philosopher. Therefore, in what follows, we provide a refutation of DPW that does not make use of assumptions that DPW is not ready to accept.

4 Logical Refutation of DPW

Without loss of generality, we will restrict our logical refutation of DPW to the logical property of being refutable (which is equivalent with being inconsistent, self-contradictory, logically false). This is usual in automated theorem proving (ATP). Since refutability and other logical properties such as theoremhood (provability, logical truth, the property of being a tautology or logical proposition) or nonrefutability (consistency, satisfiability) all depend on each other, the decision on all of these properties is reducible to the decision on one of these properties.

We take the usual syntax and semantics of FOL for granted. In particular, we consider only calculi that are correct and complete with respect to logical (in)validity as defined by the semantics of mathematical logic (model theory). This does not mean that we accept the semantic foundations of FOL; it means only that any decision procedure for deciding refutability decides a logical property that is equivalent to the canonical definition of logical invalidity. Thus, we do not allow DPW to be saved by relating it to some logic other than (traditional) FOL. The *Entscheidungsproblem* relates to traditional FOL and not to any other logic.

Let us use the more common term "normal form" instead of Wittgenstein's "adequate notation". Then, we can specify the postulate of DPW as follows. DPW postulates that the refutability of any FOL formula ϕ is decidable by converting ϕ into a normal form ψ that provides a general decision criterion for refutability by virtue of ψ's syntactic form. The conversion of ϕ into ψ is a finite, mechanical procedure that depends on the logical constants in ϕ and results in a normal form ψ that is equivalent to ϕ with respect to refutability. Instead of "syntactic form", we will also use the term "pattern". To our understanding, Wittgenstein believed in a decision procedure by means of what is currently called "pattern detection". The decision criterion in question is a pattern shared by every refutable ψ. Therefore, it can serve as a syntactic criterion for refutability. We also take for granted that WPL demands that one can read off, and thus generate, a model (a condition for truth) from ψ's pattern in the case that ϕ is satisfiable. Otherwise, the claim that ϕ/ψ shares its logical form with whatever can be represented by ϕ/ψ would hardly make sense; cf. also the quote from PR §174[4] and its discussion in section 5.

Let us state in advance that our argument is not a substitute for a *general* undecidability proof of FOL. First, we do not reduce to absurdity the *general hypothesis* of a decision procedure. Instead, we argue that the *specific* idea of DPW relying on anything similar to *known* normal forms fails for the whole realm of FOL. Second, and more importantly, we will show that our argument does not refute the general idea of deciding FOL by defining some upper bound for the application of inference rules in ATP.[4] This makes evident that our

[4] This is an idea that Gumanski [2000] and Gumanksi [2008] suggested in order to refute the Church–Turing theorem. Gumanski specified upper bounds for proofs in tableaux in terms of primitive recursive functions. Matzer [2018] and Mycka/Rosa [2018] argued independently against Gumanski's approach, asserting that his upper bound is refuted by the provability of formulas that translate into FOL functions that are known to be total and μ-recursive (and thus computable) but not *primitive* recursive functions, such as the Ackermann–Peter function. In footnote 9, we additionally show that Gumanski's definition of an upper bound is based on a misunderstanding of tableau proofs that ignores the necessity of looping. In contrast to Wittgenstein, Gumanski's attempt is motivated not by a definition of logical properties in

argument is not a substitute for a proof of the Church–Turing theorem. We argue, however, that Wittgenstein's conviction in DPW cannot be based on upper bound considerations because an upper bound for proving refutability does not ensure any pattern from which to generate a model in the case of satisfiability.

In what follows, we first briefly illustrate in section 4.1 how DPW can be realized for some fragments of FOL before we then explain in 4.2 why it cannot be realized for the whole realm of FOL.

4.1 DPW for FOL-fragments

There are many ways to realize DPW in propositional logic. Consider, e.g., disjunctive normal forms (DNFs) as normal forms ψ. Whether a propositional formula ϕ is contradictory (refutable) can be decided by the question of whether each disjunct of the DNF ψ of ϕ contains an atomic proposition A and its negation $\neg A$. This latter property is a pattern of ψ that serves as a decision criterion. If it is not satisfied, then one can directly read off a model from a disjunct that does not contain a complementary pair of atomic propositional variables.

Wittgenstein's exemplification of DPW by his *ab*-notation is similar to the case of DNFs. ϕ is contradictory iff the outermost *a*-pole is connected only to opposed innermost poles (with a symbolizing truth and b symbolizing falsehood) of one and the same propositional variable in the *ab*-diagram ψ of ϕ. If this is not the case, then the remaining connections of the outermost *a*-pole to the innermost poles of atomic propositional variables identify a model for ϕ. Truth tables provide a further exemplification of DPW in propositional logic: A line with 'T' below the main operation identifies a model, and if there is no such line, then ϕ is contradictory.

DPW is also applicable to monadic FOL (the fragment of FOL with only monadic predicates). Any monadic formula ϕ can be converted into a duplicate-free normal form ψ' in terms of a disjunction of conjunctions (monadic FOLDNF) of negated or non-negated formulas of the form $\exists\mu(A_1(\mu) \wedge \ldots \wedge A_n(\mu))$, where $n \geq 1$ and the $A_i(\mu)$ are literals with only one argument. Let us call a pair of literals, e.g. Fx and $\neg Fy$, with the same predicate, one negated and the other not negated, "opposed literals". Opposed literals that are identical except for the negation sign, e.g. Fx and $\neg Fx$, are called "complementary literals". If a conjunction in our decision procedure for monadic FOL contains a formula $\exists\mu(A_1(\mu) \wedge \ldots \wedge A_n(\mu))$ with complementary literals, then this conjunction is contradictory and can be deleted. If the resulting monadic FOLDNF is the empty word in this stage of the procedure, then ϕ is a contradiction; otherwise, the resulting conjunctions of the monadic FOLDNF can be presented as a Venn diagram ψ. In doing so, conjuncts of the form $\neg\exists\mu(A_1(\mu) \wedge \ldots \wedge A_n(\mu))$ with complementary literals are omitted because they are tautologous. If the monadic FOLDNF contains a disjunct of the form $\exists\mu(A_1(\mu) \wedge \ldots \wedge A_n(\mu))$ and a disjunct

terms of pattern detection but by a misguided, oversimplified conception of proofs in tableaux.

of the form $\neg\exists\mu(A_1(\mu) \wedge \ldots \wedge A_n(\mu))$, then all disjuncts must be deleted, and ψ is the empty word. If ψ is the empty word in this stage of the procedure, then ϕ is a tautology. ϕ is contradictory iff every disjunct of the monadic FOLDNF is converted into a Venn diagram ψ with a region that is both filled (symbolizing its emptiness) and crossed (symbolizing its nonemptiness). Otherwise, models can be read off from the regions of ψ.

Lampert [2017a] has explained how Wittgenstein's *ab*-notation can be applied to the fragment of FOL (called "elementary FOL") with formulas ϕ that can be converted into disjunctions of conjunctions in negated normal form and quantifiers pulled inwards (antiprenex formulas) without dyadic connectives in the scope of quantifiers. In this case formulas are reduced to primitive formulas (= complex poles in the *ab*-notation) with decidable logical relations (such as being contradictory or subaltern). On the basis of identifying the logical relations between primitives, one can convert all and only equivalent formulas ϕ of elementary FOL to a unique DNF ψ of primitive formulas or a unique representative ψ in the *ab*-notation in terms of a set of sets of complexe poles.

Within elementary FOL, the logical relations between primitive formulas (complex poles in the *ab*-notation) are decidable because one need not increase the complexity of the primitive formulas (complex poles in the *ab*-notation) in order to identify their logical relations. This property corresponds to the so-called finite model property, i.e. the fact that a certain fragment of FOL is decidable by considering only a finite number of models. The all important point is that a representative ψ can be generated where one can read off a finite number of possible extensions that make the initial formula ϕ true or false respectively.

The realm and results of elementary FOL can be extended to antiprenex formulas with $\wedge$ as the only dyadic connective in the scope of universal quantifiers. These formulas are known as "Herbrand formulas" and are known to be decidable. Gladstone [1966] has proven that Herbrand formulas in FOL without identity have the finite model property.

One way to generate normal forms ψ that satisfy the conditions of DPW from Herbrand formulas is by converting the Herbrand formulas into their *clause forms*. Clause forms are sets of clauses, where a clause is a set of literals. Quantifiers are eliminated by replacing existential quantifiers with Skolem functions. Since existential quantifiers are eliminated, universal quantified variables occur free. No free variable occurs in more than one clause. The approach of converting FOL formulas into clause normal form and processing the resulting clause forms is standard in ATP. Herbrand formulas can be converted into sets of clauses of length 1. In the special case of clause forms with clauses of length 1, it is sufficient to consider up to n^m substitutions of the m free variables and n Skolem functions to solve the decision problem. If none is a contradiction, then the formula is satisfiable, and a finite model with no more than n objects can be generated. Thus, any FOL formula ϕ that can be converted into a disjunction of Herbrand formulas can be reduced to a propositional DNF ψ with disjuncts that are conjunctions of the translations of all n^m substitutions of the free variables of the clause form of a Herbrand formula.

The same is true for formulas ϕ that do not have any existential quantifiers in the scope of universal quantifiers. In this case, it is also sufficient to realize n^m alternative substitutions; cf. Börger et al. [2000], Proposition 6.2.17, p. 257. Again, DPW is realizable by reducing the decision problem to a propositional DNF ψ such that we can read off a model from a disjunct that does not contain complementary literals.

However, for a case with both disjunctions and existential quantifiers in the scope of universal quantifiers, satisfiable formulas may have only infinite models. Thus, in this case, we cannot reduce the decision problem to a normal form ψ that allows one to read off models or counter-models. The question arises of how DPW treats infinity in FOL.

4.2 Infinity axioms

Without loss of generality, we will consider the problem of identifying models and countermodels for sets of clause forms. Infinity axiom sets (in short: infinity axioms) are consistent sets of clauses that have only infinite models. For the following, it will suffice to consider sets of so-called Krom–Horn clauses without identity. Krom–Horn clauses are clauses in Skolem form with maximal length 2 and at most one non-negated literal. Krom–Horn clauses correspond to formulas with disjunctions of maximal length 2 in the scope of quantifiers. Sets of Krom–Horn clauses without identity are the simplest fragment of FOL that is known to be undecidable; cf. Börger et al. [2000], section 5.1.1. Of course, we do not presume the undecidability of this FOL fragment for our reasoning. We merely refer to consistent sets of Krom–Horn clauses that have no finite models.

$$\phi: \quad \forall x_1 \neg Fx_1x_1 \wedge \forall x_2 \exists y_1 (Fx_2y_1 \wedge \forall x_3(Fx_3y_1 \vee \neg Fx_3x_2)) \tag{1}$$

$$\phi' \quad \{\{\neg Fx_1x_1\}, \{Fx_2sk_1(x_2)\}, \{Fx_3sk_1(x_4), \neg Fx_3x_4\}\} \tag{2}$$

Consider (1) and its clause form (2)[5], which is an infinity axiom set (cf. Börger et al. [2000], p. 33). Since no finite model is available, we cannot simply read off a model via a one-to-one correspondence between some finite pattern of a normal form ψ and a model.Instead, some *finite repeating* pattern must be present to identify an infinite model.[6] Since converting ϕ into a normal form

[5]Note that skolemization is followed by the rectification of universal variables. For this reason, different clauses have different variables.

[6]According to the *Tractatus*, a proposition is a truth function of atomic propositions. Truth functions are generated via the successive application of logical operations to all values of a propositional variable. TLP 5.501 distinguishes three methods for describing the values of such a variable: (i) by means of a finite enumeration, (ii) by means of a propositional function, and (iii) by means of a formal law, which we interpret as a rule for repeating a pattern. (1) is satisfiable only if there are infinitely many atomic propositions of the form Fxy, that is, if this propositional function has infinitely many values. An infinity axiom set such as (1) cannot, therefore, be translated into a finite normal form ψ in terms of a finite DNF such that a finite model can be read off from each of its disjuncts. We argue that Wittgenstein needs to stipulate normal forms ψ in terms of method (iii) in the case of infinity axioms. Such a normal form will necessarily involve the construction of "infinite conjunctions" by means of

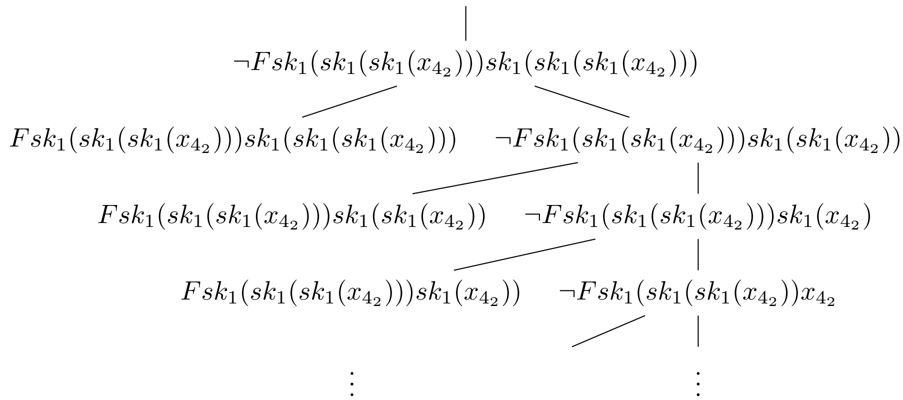

Figure 1: Endless Proof Search in ATP for (2)

ψ is an equivalence procedure with respect to refutability and ATP produces automated sequences of formulas preserving refutability, ψ must be the result of ATP.

Figure 1 presents the first four steps of a proof search in tableaux for (2). The proof search is deterministic for the special cases of Krom–Horn clauses that we consider[7], that is, there are no alternative tableaux to consider. Proof search in tableaux is identical to proof search in resolution in the special case of Krom–Horn clauses. However, our considerations are independent of the special calculus that we choose and the special normal form that we consider. Whatever calculus and normal form we choose, infinity axioms cannot be identified from a finite normal form and an ATP-search will run in an endless loop.Therefore, our question is whether it is possible to read off infinite models from some normal form ψ generated in ATP with a repeating pattern. We will discuss this for ATP in tableaux/resolution for the special case of infinity axioms in terms of Krom–Horn clauses with a deterministic proof search. From this proof search case, it will be seen that no other known proof search can do better.

The proof search for (2) in tableaux will go on forever, repeating the utilization of the same clause, with the only difference being that in the nth inference step, the variable x_{4_2} is replaced by $sk_1(x_{4_1})$ (with x_{4_1} being a new variable) if n is odd or x_{2_1} is replaced by $sk_1(x_{4_2})$ (with x_{4_2} being a new variable) if n is

method (iii), where every conjunct must ultimately refer to distinct elementary propositions in such a way that an infinite model can be read off from this construction. The question facing Wittgenstein's conjecture with respect to infinity axioms is, therefore, whether it is possible to translate a formula ϕ that describes an infinite structure into a formula ψ that contains a finite repeating pattern from which an infinite model can be read off. Cf. also our remarks at the beginning of section 5.

[7] Recall that a proof search that starts with clauses that have no positive literals is complete, cf. Letz/Stenz [2006], p. 2040. Our ATP-search for Krom-Horn clauses is based on the strong tight connection tableaux calculus described in this paper.

even. From this, it follows that the nesting of the Skolem functions increases by 1 in each inference step. This is not an exact repetition, which would allow termination of the proof search on a proof path in accordance with the regularity criterion. However, the repetition is "almost" exact in that the same inference step is repeated with the same substitutions, merely increasing the nesting of the resulting Skolem functions by 1. There is no way for any known correct calculus to make the repetitions of the inference steps and the resulting formulas more similar. This is why our considerations do not depend on a particular selection of the tableau and clause forms.

The proof search for (2) obviously enters an infinite loop, with endlessly repeating substitutions. The question for DPW is whether we can take a finite number of loops with endless repeating substitutions as a criterion for infinite models. In the following, we demonstrate that and explain why this is not the case.

$$\phi: \quad \begin{array}{c} \forall x_1 \neg F x_1 x_1 \wedge \forall x_2 \exists y_1 (F x_2 y_1 \wedge \forall x_3 (F x_3 y_1 \vee \neg F x_3 x_2) \wedge \\ \forall x_4 (F y_1 x_4 \vee \neg F_1 x_2 x_4) \wedge \forall x_5 (F_1 y_1 x_5 \vee \neg F_2 x_2 x_5) \wedge \\ \forall x_6 (F_2 y_1 x_6 \vee \neg F x_6 y_1)) \end{array} \tag{3}$$

$$\phi': \quad \begin{array}{c} \{\{\neg F x_1 x_1\}, \{F x_2 sk_1(x_2)\}, \{F x_3 sk_1(x_4), \neg F x_3 x_4\}, \\ \{F sk_1(x_5) x_6, \neg F_1 x_5 x_6\}, \\ \{F_1 sk_1(x_7) x_8, \neg F_2 x_7 x_8\}, \{F_2 sk_1(x_9) x_{10}, \neg F x_{10} sk_1(x_9)\}\} \end{array} \tag{4}$$

Consider (3) and the proof of its clause form (4) in Figure 2. (3) contains (1) as part of the expression and adds formulas of the form $\forall x_k(F_i y_1 x_k \vee \neg F_{i+1} x_2 x_k)$ (i starts with the empty word and then runs from 1 to $m-4$; k runs from 4 to $m-1$) plus a final clause $\forall x_m(F_{m-4} y_1 x_m \vee \neg F x_m y_1)$. (4) contains (2) as a proper part, and a proof of (4) must make use of all clauses in (4). Figure 2 shows that the proof starts in the same manner as the proof search in Figure 1 does; cf. the underlined literals. Through the insertion of further conjuncts of the form $\forall x_k(F_i y_1 x_k \vee \neg F_{i+1} x_2 x_k)$ into (3), the number of initial repeating steps necessary for the proof can be increased to any arbitrary number before it finally terminates by utilizing axiom 2 (= clause 2 in (4)). This demonstrates that repetitions of substitutions that do nothing but increase nesting by 1 cannot determine the occurrence of any infinite repetitions in models. They may run in an endless loop, as in Figure 1, or they may be necessary for a proof, as in Figure 2.

The critical point for DPW is that there is no pattern in the case of endless repetitions of inference steps that can serve as a sufficient criterion for the existence of infinite models and, thus, satisfiability. Of course, one can argue that the number of repetitions needed for the proof of (4) depends on the number of clauses of the form $\{F_i sk_1(x_j) x_k, \neg F_{i+1} x_j x_k\}$. However, other refutable formulas can be trivially generated in which the number of loops exceeds the number of clauses to any arbitrary extent or in which the number of loops depends on

$$\underline{\neg F sk_1(sk_1(sk_1(x_{10})))sk_1(sk_1(sk_1(x_{10})))}$$

$$\underline{F sk_1(sk_1(sk_1(x_{10})))sk_1(sk_1(sk_1(x_{10})))} \qquad \underline{\neg F sk_1(sk_1(sk_1(x_{10})))sk_1(sk_1(x_{10}))}$$

$$\underline{F sk_1(sk_1(sk_1(x_{10})))sk_1(sk_1(x_{10}))} \qquad \underline{\neg F sk_1(sk_1(sk_1(x_{10})))sk_1(x_{10})}$$

$$\underline{F sk_1(sk_1(sk_1(x_{10})))sk_1(x_{10})} \qquad \underline{\neg F sk_1(sk_1(sk_1(x_{10})))x_{10}}$$

$$F sk_1(sk_1(sk_1(x_{10})))x_{10} \qquad \neg F_1 sk_1(sk_1(x_{10}))x_{10}$$

$$F_1 sk_1(sk_1(x_{10}))x_{10} \qquad \neg F_2 sk_1(x_{10})x_{10}$$

$$F_2 sk_1(x_{10})x_{10} \qquad \neg F x_{10} sk_1(x_{10})$$

$$F x_{10} sk_1(x_{10})$$

Figure 2: Tableau Proof for (4)

the length of the literals; cf. Anonym2 [000][8].

This finding is independent of any correct and complete proof search calculus. In any case there is an infinite number of pairs $< \phi_1, \phi_2 >$ of Krom-Horn formulas such that ϕ_1 is an infinity axiom and ϕ_2 is a refutable formula such that the proof search for ϕ_1 goes on forever while the proof of ϕ_2 shares a finite number of steps with the proof search of ϕ_1. Either the calculi and the notations they are based on (e.g. with or without skolemization) produce repeating patterns in the proof search or not. In the latter case, decidability by pattern detection is impossible due to the lack of patterns. In the former case, applying the pattern of repetition as criterion of satisfiability would render the proof search incorrect by identifying refutable formulas as satisfiable.

One might still think to save DPW by stipulating that there must exist some upper bound on the number of loops with repeated substitutions that is computable from the number of syntactic features of some normal form.[9] We do not purport to show that no upper bound on the length of economical proofs is definable. However, we do argue that such reasoning cannot be motivated by WPL. Postulating that possible models are to be identified by pattern detection is well motivated by (i) decision procedures for fragments of FOL, as illustrated in section 4.1, and (ii) the philosophical project of identifying possible extensions by syntactic criteria instead of defining logical properties semantically by assignments of extensions, as argued in section 2. This postulate implies the possibility of identifying infinite extensions from repetitions of patterns. Our reasoning refutes only this implication and not the postulate of some upper bound on a proof search. The refutation of *any* attempt to decide FOL, whether or not it is reasonable or even conceivable, can be achieved only by the traditional proof of the Church–Turing theorem.

We conclude that DPW is refuted by the existence of infinity axioms and the fact that refutable formulas exist with proofs that (i) are based on all clauses of the clause forms of infinity axioms and (ii) share a finite and necessary part with the endless looping observed in the proof searches for infinity axioms. This refutes DPW because it demonstrates that there is no (finite) pattern that allows the identification of infinite models. It is, one may say, only the endless repetition itself that is indicative of infinite models. However, this is no decision criterion.

[8] This paper provides a general method of converting so called splitting Turing machines (STMs) to Krom-Horn clauses. Pairs of STMs, one non-halting and one halting, can thus be converted to Krom-Horn clauses that share the beginning of the proof search, while one loops forever and the other terminates by finding a proof.

[9] Gumanksi [2008] seems to mistakenly miss the necessity of looping on a proof branch at all. He argues for his upper bound of $\frac{(1+n)\cdot n}{2} \cdot m + n$ (m = maximum number of positions of initial literals, n = number of disjuncts of a DNF of the initial clauses) based on the assumption that in any economical proof, the number of unified positions necessary for an economical proof increases by at least 1, and argues that an upper bound is definable by the totality of all positions of all pairs of literals stemming from different initial clauses. However, this ignores the fact that to unify a pair of literals, it might be necessary to repeat the unification of one and the same pair of literals many times, where this pair of literals stems from one and the same clause.

5 Wittgenstein's Philosophical Must

One can only speculate whether Wittgenstein was aware of the existence of infinity axioms in the *Tractatus*. He did not deny an infinite domain but insisted that the number of objects was a matter that could be decided only by the application of logic (TLP 5.55, 5.557). He denied the possibility of asserting meaningful propositions about the number of objects (TLP 4.1272, 5.55). Instead, he believed that the number of objects was a presupposition of the meaning (sense) of propositions (TLP 4.1211). For him, the number of objects could only manifest through the number of names in a fully analysed language (cf. Anonym1 [000] for a recent discussion of this point). An infinite number of objects would manifest in an infinite number of names (TLP 5.535).

Therefore, Wittgenstein seemed not to necessitate a finite possibility of presenting an infinite totality by some pattern with respect to the number of objects (the cardinality of the domain). However, infinity axioms are represented not by bound variables but by FOL formulas. By definition, infinity axioms can only be satisfied in an infinite domain. By conjecturing decidability, Wittgenstein seemed to assume that, in such cases, it must be possible to translate a *description* of an infinite structure in terms of a proposition into a *presentation* of it in terms of a formal series; cf. footnote 6. According to Wittgenstein, a formal series is defined inductively and, thus, by finite means (TLP 4.1252, 4.1273, 5.2522f.). He accused Russell and Frege of having missed the possibility of describing an infinite number of propositions by a formal series (PT 5.005341) and of not distinguishing between propositional functions and operations, which prescribe how to generate a formal series through iterative application (TLP 5.25-5.252). Thus, if he was aware of infinity axioms, he might have thought that the infinity of possible models satisfying a FOL formula must be presentable by some sort of symbolic repetition. Wittgenstein argued similarly for infinite series of numbers in his middle period.[10]

In his middle period, however, he was aware of infinity axioms and seemed to have come to understand that it is unclear how his belief in pattern detection can be applied to infinity axioms. In PR §147[4], he writes:

> Each thing has one and only one predecessor; a has no successor; everything except for a has one and only one successor.' These propositions appear to describe an infinite series (and also to say that there are infinitely many things. But this would be a presupposition of the propositions' making sense). They appear to describe a *structure amorphously*. We can sketch out a structure in accordance with these propositions, which they describe unambiguously. But where can we discover this structure in them?

[10]Cf., e.g., PR §190[2]:

> I MUST [upper case added] be able to write down a part of the series, in such a way that you can recognize a law. That is to say, no description is to occur in what is written down, everything must be represented [dargestellt].

Let us translate Wittgenstein's wording into the wording of this paper. Wittgenstein considers the following set of axioms in FOL with identity:

$$\phi : \quad \begin{array}{c} \forall x \exists y (Syx \wedge \neg \exists z (Szy \wedge y \neq z)), \\ \neg \exists z Saz, \\ \forall x (x \neq a \rightarrow \exists y (Sxy \wedge \neg \exists z (Sxz \wedge y \neq z))) \end{array} \tag{5}$$

with the following intended interpretations:

$\Im_i(S(y, x)) :$ x is the successor of y.

$\Im_i(a) :$ 0.

(5) is true only in an infinite domain, e.g., the negative numbers.

The problem for Wittgenstein is that $\Im_i(\phi)$ is a *propositional description* of an infinite model ("structure") without providing any pattern presenting it. Therefore, his final question can be translated as follows: Is there some ψ, computable from ϕ, such that some infinite $\Im_i(\phi)$ can be identified from a pattern of ψ?

This is a serious question that Wittgenstein does not answer. As we saw in the previous section, it has no positive answer.

Only slightly later in PR, his philosophical dilemma becomes clear. This time, he considers not the decidability of FOL but that of equations. However, this does not make a significant difference to the principal postulate of decidability, PR §174[11,12,14]:

> Undecidability presupposes that there is, so to speak, a subterranean connection between the two sides; that the bridge cannot be made with symbols.
>
> A connection between symbols which exists but cannot be represented [dargestellt] by symbolic transformations is a thought that cannot be thought. If the connection is there, then it MUST (upper case added) be possible to see it.
>
> [...]
>
> Of course, if mathematics were the natural science of infinite extensions of which we can never have exhaustive knowledge, then a question that was in principle undecidable would certainly be conceivable.

Wittgenstein's postulate of decidability is based on a philosophical rejection of genuine propositions about infinite extensions. He seemed to believe that such a view is not reasonable and intended to refute it through philosophical analysis. *Prima facie*, this is acceptable as a philosophical *project*. However, it is in danger of immunizing itself. One must admit at some point that the whole project becomes unreasonable if it cannot cope with problems with which it must be able to cope by its own standards. One cannot simply postulate (cf. the highlighted "MUST" in several quotes in this paper) what one is willing to show; one must prove it and accept the consequences when one is not able to do so. After all, philosophical ambitions must face the reality of logic and mathematics. Denying infinite extensions while allowing for traditional FOL, which includes formulas that are neither refutable nor satisfiable within a finite domain, is not compatible. The former forces one to restrict oneself to decidable fragments of FOL with the finite model property for philosophical reasons, which is not acceptable from a logical point of view taking full FOL as standard logic.

Abbreviations of Wittgenstein's Writings

CL: *Cambridge Letters*, Oxford: Blackwell.

NB: *Notebooks 1914-1916*, Oxford, Blackwell, 1979.

NL: "Notes on Logic", in *Notebooks 1914-1916*, Oxford, Blackwell, 1979, 93-107.

PR: *Philosophical Remarks*, Oxford: Blackwell, 1975.

PI: *Philosophical Investigations*, 4th edition, Indeanapolis: Wiley Blackwell, 2009.

PT: *Prototractatus*, London: Routledge, 1971.

RFM: *Remarks on the Foundations of Mathematics*, 2nd edition, Cambridge/ Ms: M.I.T. Press, 1967.

TLP: *Tractatus Logico-Philosophicus*, London: Routledge, 1994.

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

References

XXX (000).

YYY (000).

ZZZ (000).

Anscombe, E. (1959): *An Introduction to Wittgenstein's Tractatus*, London Hutchinson University Library.

Baatz, M., Egly, U., Leitsch, A. (2006): "Normal Form Transformations", in Robinson, A., Voronkov, A. (eds.): *Handbook of Automated Reasoning* I, Amsterdam: Elsevier, 273-334.

Black, M. (1964): *A Companion to Wittgenstein's Tractatus*, Cambridge: Cambridge University Press.

Böerger, E., Grädel, E., Gurevich, Y. (2000): *The Classical Decision Problem*, Berlin: Springer.

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

Church, A. (1936): "A Note on the Entscheidungsproblem", *The Journal of Symbolic Logic* 1, 40-41, 101-102.

Etchemendy, J. (1990): *The Concept of Logical Consequence*, Cambridge/MA: Harvard University Press.

Fogelin, R.J. (1976): *Wittgenstein*, London: Routledge.

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

Hähnle, R. (2006): "Tableaux and Related Methods", in Robinson, A., Voronkov, A. (eds.): *Handbook of Automated Reasoning* I, Amsterdam: Elsevier, 101-178.

Kripke, S. (2013): "The Church-Turing 'Thesis' as a Special Corollary of Gödel's Completeness Theorem," in B. J. Copeland, C. Posy, and O. Shagrir (eds.): *Computability: Turing, Gödel, Church, and Beyond*, Cambridge/Ms: MIT Press, 77-104.

Gladstone, M.: "Finite Models for inequalities", *Journal of Symbolic Logic* 31, 581-592.

Gumański, L. (2000): "The Decidability of the First-Order Functional Calculus", *Ruch Filosoficzny* 57.3, 411-438.

Gumański, L. (2008): "A New Proof of Decidability of the First-Order Functional Calculus", *Ruch Filosoficzny* 65.3, 419-438.

Lampert, T. (2017): "Wittgenstein's *ab*-notation: An Iconic Proof Procedure", *History and Philosophy of Logic* 38.3, 239-262.

Lampert, T. and Säbel, M (2021): "Wittgenstein's Elimination of Identity for Quantifier-Free Logic", *Review of Symbolic Logic* 14.1, 1-21.

Landini, G. (2007): *Wittgenstein's Apprenticeship with Russell*, Cambridge: Cambridge University Press.

Landini, G. (2021): "Showing in Wittgenstein's ab-Notation," in Wuppuluri, S., de Costa, N. (eds.): *Wittgenstein: Looking at the World from the Viewpoint of Wittgenstein's Philosophy*, Wien: Springer, 193-226.

Letz, R., Stenz, G.: "Model Elimination and Connection Tableau Procedures" in Robinson, A., Voronkov, A. (eds.): *Handbook of Automated Reasoning* II, Amsterdam: Elsevier, 2015-2114.

Matzer, M.: *Die Unentscheidbarkeit der Quantorenlogik. Widerlegung des Vorschlags für ein Entscheidungsverfahren von Leon Gumanski*, unpublished manuscript, 2018.

Mycka, J., Rosa, Wojciech (2018): "Związki problemu nierozstrzygalności logiki predykatów pierwszego rzedu z zagadnieniem złożoności", in Woleński J., Murawski, R. (eds.): *Problemy filozofii matematyki i informatyki*, Poznań: Wydawnictwo Naukowe UAM, 191-204.

Potter, M. (2009): *Wittgenstein's Notes on Logic*, Oxford: Oxford University Press.

Ray, G. (1996), "Logical Consequence: A Defence of Tarski", *Journal of Philosophical Logic* 25, 617-677.

Rogers, B. and Wehmeier, K. (2012): "Tractarian first-order logic: Identity and the N-operator", *Review of Symbolic Logic* 5.4, 538-573.

Shapiro, S. (1998), "Logical Consequence: Models and Modality", in Schirn. M. (ed.): *Philosophy Of Mathematics Today*, Oxford: Oxford University Press, 131-156.

Sundholm, G. (1990): "Sätze der Logik: An Alternative Conception", in Haller, R. (ed.): *Wittgenstein - Eine Neubewertung. Akten des 14. Internationalen Wittgenstein Symposiums 3*, Wien: Hölder-Pichler-Tempsky, 59-61.

Turing, A. (1936/37): "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society* 2.42, 230-65.

Weiss, M. (2017): "Logic in the Tractatus", *Review of Symbolic Logic* 10.1, 1-50.