Proving that the Mind is Not a Machine?*

Johannes Stern

Department of Philosophy

University of Bristol

johannes.stern@bristol.ac.uk

Abstract

This piece continues the tradition of arguments by John Lucas, Roger Penrose and others to the effect that the human mind is not a machine. Kurt Gödel thought that the intensional paradoxes stand in the way of proving that the mind is not a machine. According to Gödel, a successful proof that the mind is not a machine would require a solution to the intensional paradoxes. We provide what might seem to be a partial vindication of Gödel and show that if a particular solution to the intensional paradoxes is adopted, one can indeed give an argument to the effect that the mind is not a machine.

1 Introduction

It has seemed to a number of prominent philosophers and scientists that Gödel's Incompleteness theorems "*prove that Mechanism is false, that is, that minds cannot be explained as machines*" (Lucas, 1961, p. 112). Unfortunately, so far little evidence has been produced in support of this sentiment. The proofs that purportedly show that the mind cannot be a machine have

*This is a pre-peer-reviewed version of the article published in *Thought: A Journal of Philosophy*,7(2):81-90, 2018. Changes to the printed article are minor, however.

not been generally accepted. Rather, these so-called proofs have been widely found to be unsound and falling short of establishing that Mechanism is false. In this piece, we propose a new such "proof": if one adopts a particular formalization of Mechanism and a specific solution to the semantic and intensional paradoxes, one can show that the mind is not a machine.

The basic idea of these Gödelian arguments against Mechanism is that Gödel's Incompleteness theorems imply that no formal system, that is no machine, can prove all mathematically true sentences.[1] The Incompleteness theorems tell us that for each formal system there will always be true sentences, for example the Gödel sentence of the system, which the formal system cannot prove. However, by following the reasoning of the proof of Gödel's First Incompleteness theorem the human mind, so the argument usually goes, can establish for any given formal system that the Gödel sentence of the system is true. If this reasoning were correct, we could conclude that the theorems the human mind can prove cannot be produced by a machine. Yet, as has been pointed out by a number of authors, the Gödel sentence of a formal system is true, only if the system is consistent. So, to establish that the Gödel sentence of a system is true, the human mind needs to be able to establish the consistency of this formal system. In order to complete the argument, the Anti-Mechanist thus has to argue that the human mind can establish or perceive the consistency of any sound formal system. But this is a very strong assumption, arguably too strong an assumption, even if we impose strong idealization conditions on the human mind as it is usually done. So, one would expect the proponents of the Gödelian arguments against Mechanism to provide convincing arguments for this assumption. At least the first generation of Gödelian arguments by Lucas (1961) and Penrose (1989) fail to provide such independent arguments.[2] Moreover, it was convincingly argued that without additional assumptions there is very little hope for such an Gödelian argument against Mechanism to succeed: in a very natural formal framework it is consistent to maintain the Mechanistic thesis: it

[1] Throughout this paper we identify "machine" with "Turing machine". Authors such as Copeland (2000) reject this identification and adopt a wider formulation of Mechanism. But in this paper we focus on Gödelian arguments against Mechanism along the lines of Lucas (1961) and Penrose (1989, 1994, 1996) and these arguments can be successful only, if this identification is accepted.

[2] See, for instance, Benacerraf (1967), Shapiro (2003), or Koellner (2016, 2018a) for discussion.

can not be refuted.[3]

Gödel (1995) himself did not think that his incompleteness theorems imply that the mind is not a machine. Rather he thought that the incompleteness theorems imply that "*(…) either the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems (…)*". This disjunctive conclusion is also known as Gödel's Disjunction. Gödel's Disjunction is actually a conditional claim: it asserts that if the human mind is a machine, then there are sentences the human mind can neither prove nor refute, that is, there are absolutely undecidable sentences. If we accept Gödel's terminology his disjunctive conclusion indeed follows from the Incompleteness theorems (Koellner, 2016, 2018a,b). But, as Gödel argued and we have also pointed out, this by no means establishes the first disjunct, namely, that "*the human mind infinitely surpasses the powers of any finite machine*".

Originating with Gödel's discussion, the term *absolute provability* was used in contrast to formal provability or provability relative to a formal system. Absolute provability is supposed to be an intuitive notion of provability that is meant to capture the process by which the human mind produces mathematical theorems. With this terminology in place the question of whether the mind is a machine amounts to the question of whether the absolutely provable sentences can be generated by an effective algorithm. From our previous discussion, it ought to be clear that without further assumptions about absolute provability there is very little hope of providing a successful refutation of Mechanism. Indeed, in his *New Argument* Penrose (1994, 1996) introduced further such assumptions about absolute provability.[4] In contrast to the older generation of arguments, which are, for the most part, simply question begging, Penrose's *New Argument* fails to refute Mechanism for more subtle reasons: the prima facie plausible assumptions on behalf of absolute provability turn out to be jointly inconsistent, that is, they lead straight into *intensional paradox*.[5]

[3]These results are due to Reinhardt (1986a) and Carlson (2000). See, for instance, Koellner (2016, 2018a) for a discussion.

[4]See Shapiro (2003),Lindström (2001, 2006) and Koellner (2016, 2018a,b) for a use of this label.

[5]See, e.g., Chalmers (1995); Shapiro (2003); Koellner (2016, 2018a,b) and Stern (2017) for discussion.

According to Wang (1996), Gödel thought that *"if one could clear up the intensional paradoxes somehow, one would get a clear proof that mind is not a machine"* (Wang, 1996, p. 187). Indeed, we will show that if a specific "solution" to the intensional paradoxes is adopted, one can give a proof that the mind is not a machine given a suitable formalization of this claim, albeit not along the lines of Penrose's original argument. The "solution" to the intensional paradoxes we have in mind connects the intensional to the semantic paradoxes and assumes the paradoxes of truth to be the source of all these paradoxes. As a consequence of this strategy, if a "solution" to the paradoxes of truth in the form of a consistent theory of truth is provided, the intensional paradoxes disappear alongside. As it turns out, if with Maudlin (2004) the truth theory KF (sometimes also called FM) is chosen to be the appropriate solution to the paradoxes of truth, then we can prove that the mind is not a machine, given a suitable formalization of this claim.[6]

2 Truth and Intensional Paradox

Work by Myhill (1960) and Montague (1963) has taught us that if we treat modal and intensional notions as predicates paradox will arise. This holds in particular for the notion of absolute provability. The two constitutive principles of absolute provability, formalized by a sentential predicate K,

(T)	$K^{\ulcorner}\phi^{\urcorner} \rightarrow \phi$
(Nec)	if ϕ is a theorem, then so is $K^{\ulcorner}\phi^{\urcorner}$

for all sentences ϕ of the language, are jointly inconsistent.[7] Our proposed resolution of the paradox is to explicitly introduce a truth predicate. The idea is that (T) and (Nec) implicitly

[6]The theory KF (*Kripke-Feferman*) was developed by Feferman, cf. Feferman (1991), and vigorously defended by Maudlin (2004). However, the first written presentation and discussion of the theory was in Reinhardt (1986b). In Field (2008) the theory is called FM (*Feferman-Maudlin*).

[7]$^{\ulcorner}\phi^{\urcorner}$ is a name, e.g. the numeral of the Gödel number, of the sentence ϕ.

assume a naive truth predicate, which in the presence of self-referential sentences leads to paradox, as Gödel and Tarski have taught us. As a consequence, the naive truth predicate has to be replaced by a non-naive truth predicate for which ϕ and $T^{\ulcorner}\phi^{\urcorner}$ are no longer equivalent or intersubstitutable in all contexts. This means making the truth predicate explicit in formulating the principles of absolute provability. We are led to the following alternative principles of absolute provability, which at least prima facie seem to enjoy the same intuitive support as (T) and (Nec):

($\mathsf{T_K}$)	$\forall x(Kx \rightarrow Tx)$
(T-Nec)	if $T^{\ulcorner}\phi^{\urcorner}$ is a theorem, then so is $K^{\ulcorner}\phi^{\urcorner}$.

Whether paradox will arise now only depends on the theory of truth we adopt. If this theory is consistent, then the theory of truth and absolute provability will also be consistent: no intensional paradox will arise.[8]

In three recent papers, Koellner (2016, 2018a,b) uses this strategy for resolving the intensional paradoxes to investigate arguments against Mechanism. Koellner raises the general challenge for the Anti-Mechanist to provide a philosophically motivated theory of truth in which a convincing argument for the claim that the mind is not a machine can be carried out. Koellner remains skeptical in this regard and provides us with good reasons for his skepticism. He shows that if Feferman's (2008) attractive theory of truth DT (*Determinate Truth*) is adopted, then no argument against Mechanism can be successful: Mechanism is consistent with the theory DT supplemented by plausible principles for absolute provability that are formulated following the strategy outlined above. Indeed, Koellner shows that that the Mechanistic thesis is independent: it can be neither proved nor refuted.

Inspired by Koellner's challenge we investigate the consequences of adopting the theory of truth KF, instead of the theory DT, as the solution of the paradoxes of truth. KF may be

[8]The strategy of resolving the intensional paradoxes by connecting the intensional notions to a non-naive notion of truth has been independently developed by Stern (2014a,b, 2016) and Koellner (2016, 2018b).

viewed as an axiomatization of Kripke's theory of truth in classical logic and is perhaps the most popular theory amongst the classical theories of truth. In particular, Maudlin (2004) has defended KF as the correct solution of the paradoxes of truth. The theory is in an important sense compositional: the truth predicate commutes with all logical connectives and quantifiers with the exception of negation. For the purpose of this paper we will only rely on two aspects of KF.[9] First, we need the truth predicate to distribute over a disjunction:

$$(\vee\mathrm{D}) \qquad T^{\ulcorner}\phi \vee \psi^{\urcorner} \rightarrow T^{\ulcorner}\phi^{\urcorner} \vee T^{\ulcorner}\psi^{\urcorner} \qquad \text{for all } \phi, \psi.$$

Second, we need to use the fact that KF proves one direction of the T-scheme:

$$(T\text{-Out}) \qquad T^{\ulcorner}\phi^{\urcorner} \rightarrow \phi \qquad \text{for all } \phi.$$

Let us call the theory extending Peano Arithmetic by the principles $(\vee\mathrm{D})$ and $(T\text{-Out})$, together with $(\mathsf{T_K})$, APT (*Absolute Provability and Truth*). APT is a consistent theory.[10] As we shall see, at least under a particular formalization of Mechanism, APT proves that the mind is not a machine.

3 The Argument

In the framework of the Lucas-Penrose arguments Mechanism is the thesis that the mathematical theorems the idealized human mind can prove can be generated by some effective procedure. If we identify "machine" with "Turing machine" this means, assuming Church's thesis, that the theorems the idealized human mind can produce are the output of a Turing

[9] For further in depth discussion of KF we refer the reader to Halbach (2011).

[10] The consistency of APT is a direct corollary of the consistency KF. APT can be interpreted in KF by translating both 'K' and 'T" by the truth predicate of KF. The reader may wonder why the rule $(T\text{-Nec})$ is not included in our formulation of APT: we have omitted the rule since it is not required for carrying out the argument to the effect that the mind is not a machine.

machine.[11] Availing ourselves to the notion of absolute provability, this means that the set of absolutely provable sentences can be recursively enumerated. That is, there exists an explicit system of axioms and rules that proves all the absolutely provable sentences. Accordingly, to refute mechanism one has to show that there exists no explicit system of axioms and rules that proves all the absolutely provable sentences. This would establish that the mind cannot be a machine, that is, Anti-Mechanism. Assuming that there was an explicit system of axioms and rules that proves all the absolutely provable sentences, can we say anything about what kind of system this would be? This depends on what status we attribute to the underlying logical laws. If we ascribe to the logical principles a special status in contrast to the remaining axioms and rules, then we may assume that each such explicit system of axioms and rules proves all logical truth independent of which axioms and rules we ultimately adopt. This means that if we take classical logic as our logic in question then all classical logical truths will be provable in the system under consideration, no matter which system it is. Let us call this position *Orthodoxy about Logic*. Orthodoxy about Logic need not be accepted. It is perfectly conceivable that logic does not have a special status in comparison to the remaining theoretical postulates and, more to the point, that we may not assume at the outset that the logical truths of one particular logic will be provable in each of the formal systems considered. In this case the only assumption we are allowed to make is that the theorems of the explicit system of axioms and rules are recursively enumerable, i.e., are the output of a Turing machine. This is a more general formulation of Mechanism and it is the one Koellner (2016, 2018a,b) adopts in his discussion.This would be some form of Non-Orthodoxy about Logic. While Orthodoxy about Logic is by no means generally accepted, it underlies, at least implicitly, a good number of philosophical discussions and it is thus not unreasonable to see where it leads us in the context of the present debate—especially since a classical solution to the paradoxes has been adopted. Consequently, in what is to come we will take such an orthodox stance and will consider classical logic to be sacrosanct and not up for debate.

[11] As we noted earlier this identification need not be accepted. In this case the Mechanistic thesis might take a very different shape. See Copeland (2000) for discussion.

Under these assumptions Mechanism is equivalent to the claim that the absolutely provable sentences can be proved in classical logic from a recursive set of sentences and, assuming this characterization of Mechanism, we now move on to turning Mechanism and Anti-Mechanism into precise formal claims. To this end we use the predicate K for absolute provability again. Moreover, if Σ is a recursive set of axioms of some theory $\mathcal{T}$, we let σ be a natural representation of this set in a language extending the arithmetical language of, say, Peano Arithmetic. Let Pr_σ be a natural provability predicate of $\mathcal{T}$. Given these stipulations Mechanism amounts to the following thesis:

MEC $$\exists\sigma\forall x(Kx \leftrightarrow \mathsf{Pr}_\sigma(x)).$$

A refutation of Mechanism would reject this claim. That is, Anti-Mechanism would be the thesis that there is no recursive set of sentences Σ from which all absolutely provable sentences follow:

ANTIMEC $$\neg\exists\sigma\forall x(Kx \leftrightarrow \mathsf{Pr}_\sigma(x)).$$

Following the outlines of the traditional arguments by Lucas and Penrose, our refutation of Mechanism will proceed via a reductio strategy: we will assume that the absolutely provable sentences coincide with the theorems of some recursively axiomatizable theory $\mathcal{T}$ and we will derive a contradiction starting from this assumption. That is, we will assume

$$\forall x(Kx \leftrightarrow \mathsf{Pr}_\sigma(x))$$

for some σ. However, it is important to notice that throughout the reductio proof we may not assume that the human mind knows *which* recursively axiomatizable theory it is—this would not amount to a refutation of Mechanism but of a much stronger claim.[12] In particular, even

[12]See Benacerraf (1967); Reinhardt (1986a) and Shapiro (2003) for further discussions.

though the reductio argument will be carried out in APT and, implicitly, we may assume that in our reductio assumption Pr_σ stands for the provability predicate of APT, we may not infer $\mathsf{Pr}_\sigma(\ulcorner \phi \urcorner)$ whenever we have proved ϕ.[13]

Before we give the argument, we introduce the standard Liar sentence, that is, a sentence λ such that APT, or any other arithmetical theory extending Q in an arithmetical language containing the truth predicate, proves:

$$(L) \qquad \neg T\ulcorner \lambda \urcorner \leftrightarrow \lambda.$$

We can now give the argument to the effect that the mind is not a machine. We reason in APT:

Assume for reductio that the mind is a machine.

$$(*) \qquad \forall x(K(x) \leftrightarrow \mathsf{Pr}_\sigma(x))$$

By the principle $(\mathsf{T_K})$ the reductio assumption implies

$$(1) \qquad \forall y(\mathsf{Pr}_\sigma(y) \rightarrow T(y))$$

and by universal instantiation

$$(2) \qquad \mathsf{Pr}_\sigma(\ulcorner \lambda \vee \neg\lambda \urcorner) \rightarrow T(\ulcorner \lambda \vee \neg\lambda \urcorner).$$

Since $\lambda \vee \neg\lambda$ is a classical tautology it is provable *independently* of which axioms are assumed. Therefore $\lambda \vee \neg\lambda$ is provable relative to *any* set of axioms and, in particular, it is provable

[13] From the more technical point of view this means that we cannot appeal to Löb's derivability conditions when reasoning about Pr_σ in the reductio proof.

relative to the set of axioms at stake:

(3) $$\mathsf{Pr}_\sigma(\ulcorner \lambda \vee \neg\lambda \urcorner).$$

By (2) this yields

(4) $$T\ulcorner \lambda \vee \neg\lambda \urcorner.$$

Due to $(\vee\mathrm{D})$ the truth predicate commutes with disjunction and hence

(5) $$T\ulcorner \lambda \urcorner \vee T\ulcorner \neg\lambda \urcorner.$$

Because of (L) the left disjunct of (5) is equivalent to $\neg\lambda$. But due to $(T\text{-}\mathsf{Out})$ the right disjunct also implies $\neg\lambda$. We can infer

(6) $$\neg\lambda.$$

By (L) the latter implies $T\ulcorner \lambda \urcorner$ and thus by $(T\text{-}\mathsf{Out})$ we derive

(7) $$\lambda.$$

This ends the reductio proof since it contradicts (6). We conclude

$$\neg\forall x(K(x) \leftrightarrow \mathsf{Pr}_\sigma(x)).$$

Moreover, we have not introduced any assumption concerning σ and therefore we can introduce the universal quantifier

$$\forall\sigma\neg\forall x(K(x) \leftrightarrow \mathsf{Pr}_\sigma(x)).$$

The latter is clearly equivalent to AntiMec: the mind is not a machine. □

The argument exploits the fact that in KF-style theories of truth and absolute provability the so-called internal logic, that is the logic within the scope of the truth predicate, and the external logic diverge. The external logic is just classical logic while the internal logic of KF-style theories is strong Kleene logic. As a consequence, classical tautologies are not generally true in the object-linguistic sense and cannot be for sake of consistency. However, we have assumed classical logic throughout and, in particular, our reductio assumption, i.e. $\forall x(K(x) \leftrightarrow \mathsf{Pr}_\sigma(x))$, reflects this fact since the axioms of classical logic are built into the standard provability predicate. This implies that we are only considering theories formulated in classical logic and thus no matter which theory $\mathcal{T}$ we consider, $\mathcal{T}$ will prove the classical tautologies. As a consequence (T_K) has the effect of adding the classical tautologies to the internal logic of the theory. But, as our argument shows, this leads to a contradiction in KF-style theories of truth and absolute provability and we may conclude that no recursively axiomatizable theory $\mathcal{T}$ can produce the same theorems as the human mind. The mind cannot be mechanized.

However, in APT we cannot show that Mec is *false*, that is, we cannot show that it is *true* that the mind is not a machine. In the terminology of Kripke (1975) Mec is an ungrounded sentence and, as a consequence, neither Mec nor AntiMec can be proved to be true in APT or, more generally, KF-style theories of truth and absolute provability. The best we can say is that Mec is *not true* but this does not imply that it is *false* since the latter would imply the truth of AntiMec. This is just one example of the general phenomenon that in KF-style theories provability and truth *provably* come apart: in such theories *provability outstrips truth*.

4 The Moral

Let us assume for the moment that Orthodoxy about Logic is correct. Then the strength of the argument clearly depends on whether we take KF-style theories of truth to be an acceptable solution to the paradoxes of truth. In the previous section, we have seen several features of these

theories that suggest that this might not be without problems. The main problem is perhaps that according to their own standard, KF-style theories of truth are not to be trusted: there are sentences that the theory proves, in fact that all classical recursively axiomatizable theories of the language prove, and which are provably not true. However, despite these problems there have been advocates of KF in the literature. Most notably, Maudlin (2004) is a heroic philosophical defence of KF qua theory of truth and solution to the paradoxes of truth. Maudlin supplements KF with particular norms for assertion and denial that are designed to account for the clash between provability and truth in KF. Unsurprisingly, according to these norms *assertability outstrips truth*: asserting a sentence does not generally commit one to the truth of the sentence. It is *permissible* to assert some sentences that are neither true nor false, namely the consequences of the semantic theory, that is KF. For these *permissible* sentences, it is not only *permissible* to assert them but we can even *believe* (and *defend*) these sentences as long as we don't believe them to be true. However, other sentences that are neither true nor false are *impermissible* in that sense. For example, it would be *impermissible* according to Maudlin's norms for assertion and denial to assert the negation of a theorem of the theory. This is important because it shows that even though it is possible to assert sentences that are *not true* this is not tantamount to trivializing the norms for assertion and denial.

Let us assume Maudlin is right and accept KF as the solution to the paradoxes of truth. What does our argument then show? It shows that Mec is impermissible and *not true*. Under no circumstances can we assert Mechanism or believe it. In contrast, it is permissible to deny Mec and, equivalently, to assert Anti-Mechanism. We are even licensed to believe this claim. However, we may not assert or believe that AntiMec is true. Even though KF does not license this final step in the refutation of Mechanism our argument provides a sound argument for Anti-Mechanism if Orthodoxy about Logic is assumed. In this sense, it supports philosophers and scientists in their belief that Gödel's Incompleteness theorems show—albeit in a very indirect way— that minds cannot be explained as machines as long as they do not equivocate it to the belief that Mechanism is false. But what will the Mechanist make of the argument and

its startling conclusion?

Acknowledgment This work was supported by the European Commission through a Marie Sklodowska-Curie Individual Fellowship (TREPISTEME, Grant No. 703539). I wish to thank Catrin Campbell-Moore, Leon Horsten, Carlo Nicolai and especially Peter Koellner for helpful comments.

References

Benacerraf, P. (1967). God, the Devil, and Gödel. *The Monist*, 51(1):9–32.

Carlson, T. J. (2000). Knowledge, machines, and the consistency of reinhard's strong mechanistic thesis. *Annals of Pure and Applied Logic*, 105:51–82.

Chalmers, D. J. (1995). Mind, Machines, and Mathematics. a Review of Shadows of the Mind by Roger Penrose. *Psyche*, 2(9).

Copeland, B. J. (2000). Narrow Versus Wide Mechanism: Including a Re-Examination of Turing's Views on the Mind-Machine Issue. *Journal of Philosophy*, 79(1):5–32.

Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic*, 56:1–47.

Feferman, S. (2008). Axioms for determinateness and truth. *Review of Symbolic Logic*, 1:204–217.

Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press.

Gödel, K. (1995). Some basic theorems on the foundations of mathematics and their implications. In Feferman, S., Dawson Jr., J. W., Goldfarb, W., Parsons, C., and Solovay, R. M., editors, *Kurt Gödel: Collected Works*, volume 3, pages 304–323. Oxford University Press, Oxford. Manuscript written in 1951.

Halbach, V. (2011). *Axiomatic Theories of Truth*. Cambridge University Press.

Koellner, P. (2016). Gödel's disjunction. In Horsten, L. and Welch, P., editors, *Gödel's Disjunction. The Scope and Limits of Arithmetical Knowledge*, pages 148–188. OUP.

Koellner, P. (2018a). On the Question of Whether the Mind can be Mechanized, I: From Gödel to Penrose. *The Journal of Philosophy*, to appear.

Koellner, P. (2018b). On the Question of Whether the Mind can be Mechanized, II: Penrose's New Argument. *The Journal of Philosophy*, to appear.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, 72:690–716.

Lindström, P. (2001). Penrose's New Argument. *Journal of Philosophical Logic*, 30:241–250.

Lindström, P. (2006). Remarks on Penrose's "New Argument". *Journal of Philosophical Logic*, 35:231–235.

Lucas, J. R. (1961). Minds, Machines, and Gödel. *Philosophy*, 36:112–127.

Maudlin, T. (2004). *Truth and Paradox. Solving the Riddles.* Oxford University Press.

Montague, R. (1963). Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica*, 16:153–167.

Myhill, J. (1960). Some remarks on the notion of proof. *The Journal of Philosophy*, 57:461–471.

Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* OUP.

Penrose, R. (1994). *Shadows of the Mind: In Search for the Missing Science of Consciousness.* OUP.

Penrose, R. (1996). Beyond the Doubting of a Shadow. *Psyche*, 2(23).

Reinhardt, W. N. (1986a). Epistemic Theories and the Interpretation of Gödel's Incompleteness Theorems. *The Journal of Philosophical Logic*, 15(4):427–474.

Reinhardt, W. N. (1986b). Some remarks on extending and interpreting theories with a partial predicate for truth. *The Journal of Philosophical Logic*, 15:219–251.

Shapiro, S. (2003). Mechanism, Truth, and Penrose's New Argument. *Journal of Philosophical Logic*, 32:19–42.

Stern, J. (2014a). Modality and Axiomatic Theories of Truth I: Friedman-Sheard. *The Review of Symbolic Logic*, 7(2):273–298.

Stern, J. (2014b). Modality and Axiomatic Theories of Truth II: Kripke-Feferman. *The Review of Symbolic Logic*, 7(2):299–318.

Stern, J. (2016). *Toward Predicate Approaches to Modality*, volume 44 of *Trends in Logic*. Springer, Switzerland.

Stern, J. (2017). Penrose's *New Argument* and paradox. In Piazza, M. and Pulcini, G., editors, *Truth, Existence, and Explanation. FilMat Studies in the Philosophy of Mathematics*, Boston Studies in the History and Philosophy of Science. Springer. forthcoming.

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