Penultimate draft

Forthcoming in: *Studia Semiotyczne (Semiotic Studies)* (a special issue on the Gödelian anti-mechanist arguments, edited by Roman Kossak)

Remarks on the Gödelian Anti-Mechanist Arguments

Panu Raatikainen

Introduction

There is no question that Gödel's two incompleteness theorems (Gödel 1931)[1] are deep and important mathematical results which have significant philosophical implications.[2] It seems that the idea that they demonstrate the superiority of the human mind over computing machines and formalized theories in particular is very attractive and natural, as it is put forward again and again (see Krajewski, this volume). The view that the human mind is, in some sense, equivalent to a finite computing machine or a formalized theory is called "mechanism". The popular idea, famously advocated by J.R. Lucas (1961, 1996), is that Gödel's results demonstrate, with mathematical certainty, that the human mind can surpass or "out-Gödel" any computing machine and formalized theory, and that mechanism can therefore be refuted for good. Roger Penrose (1989, 1994, 1995, 1997) has prominently put forward very similar views. The literature is vast, but I shall focus here on Lucas's classic key claims, which have been enthusiastically repeated, almost verbatim, again and again.[3]

However, numerous logicians and philosophers (beginning from Gödel and Turing themselves; and see e.g. Putnam 1960; Boolos 1968; Davis 1990a, 1990b; Feferman 1995, 2009, 2011; Shapiro 1998) including myself (Raatikainen 2005) have argued that such straightforward anti-mechanist arguments grounded on Gödel's theorems are flawed. Krajewski (this volume) both surveys the history of such arguments and elaborates various problems with them. I don't want to repeat those critical arguments here in any detail. Instead, I shall emphasize and discuss certain selected issues around the Gödelian anti-mechanist arguments which have received less attention, and which to my mind deserve to be noticed. I shall assume that the reader is familiar with the basic ideas and concepts of this debate.

[1] For an accessible survey, see e.g. Franzén 2005, Raatikainen 2020.

[2] See e.g. Raatikainen 2005.

[3] There is a conspicuously enthusiastic entry (Megill 2012) on the Gödelian anti-mechanistic arguments in *the Internet Encyclopedia of Philosophy*, which for its part suggests that the issue is still very much alive.

The Limits of Machine Talk

The Gödelian argument against mechanism is standardly formulated in terms of Turing machines and their Gödel sentences, which the machines are incapable of "producing as being true" but which the human can allegedly see to be true. However, such talk about "the Gödel sentence of the machine" is, strictly speaking, nonsense (cf. Gaifman 2000).

The theory of computability and its notions of decidability and computability, and Turing's groundbreaking analysis of these notions in terms of imaginary idealized machines, are certainly essential for the general versions of the incompleteness results: a formalized theory is by definition required to have a decidable set of axioms and a decidable proof relation.[4] Consequently, if the language of the theory is suitably coded by numbers, there are Turing machines which can effectively generate exactly the code-numbers (the "Gödel numbers") of the theorems of the theory: the set of those code-numbers is thus, technically speaking, recursively enumerable (r.e. for short). But that is it.

A Turing machine does not in itself correspond to any specific formalized theory, and just does not have a specific Gödel sentence of its own. Even if a Turing machine is incapable of producing ("as true") a sentence under one given coding, it may well produce that sentence under many other codings. Consequently, the suggested idea of "out-Gödeling" a machine in itself makes no sense. One and the same Turing machine may correspond to very different formalized theories under different codings. And many Turing machines just do not correspond to any formalized theory under any coding. The same holds for recursively enumerable sets of numbers.

The framework of computability theory is, in general, too coarse-grained in this context: all formalized theories which contain Robinson Arithmetic **Q**, from the very weak **Q** itself to the strongest theories of set theory (e.g. **ZFC** + "there exist supercompact cardinals") and beyond, as long as the set of axioms is decidable, are "creative" (in Post's sense), have the same computability-theoretic degree, and are recursively isomorphic with each other (i.e., they are all one-one reducible to each other). Hence computability theory is unable to make any difference between such theories with radically different strengths. As Kreisel was fond of putting it, proof theory begins where computability theory stops. (cf. Odifreddi 1989, p. 356–7) Hence, immaculately formulated, the question should be: Can all the truths that are humanly provable be captured by a formalized theory?

My sticking to this issue may strike some as excessive pedantry, but I think there is a real risk here of overlooking some relevant issues. Turing machines, or r.e. sets of numbers generated by them, simply do not stand to each other in the various logical relations that are essential for this topic. For example, it makes no clear sense to ask whether a given r.e. set of numbers can prove the consistency of another given r.e. set. Furthermore, in a fully general consideration, we cannot restrict our attention solely to direct extensions of elementary arithmetic in the same language (or its direct extension). A great many formalized theories have *prima facie* nothing to do with arithmetic; their language may be

[4] See e.g. Raatikainen 2020. To be sure, logicians have studied extensively arbitrary sets of axioms, infinitely long sentences and infinitary rules of inference; but in the context of Gödel's incompleteness theorems, this is a standard assumption (though some generalizations exist). Accordingly, in what follows, I shall always use "formalized theory" to mean a theory which has a finite or decidable set of axioms, and a decidable proof relation, and consequently an r.e. set of theorems.

quite different from the familiar language of arithmetic (think of set theory, for example). There is no direct way of comparing the respective sets of code-numbers, as to whether one is stronger than the other, etc. Relating such theories requires considering the relation of *relative interpretability* between formalized theories.[5] But at the level of computability theory and r.e. sets, such relations are invisible.

It is certainly possible to continue to talk about Turing machines or recursively enumerable sets of numbers here, with the assumption that some coding ("Gödel numbering") has been fixed. But such a manner of speaking may be misleading and hide some important aspects of the topic. The above facts should at very least be kept clearly in mind. Accordingly, I shall talk, in what follows, as far as possible, only about formalized theories.

Varieties of Mechanist and Anti-Mechanist Theses

Instead of talking generally about the juxtaposition of mechanism and anti-mechanism, I think it would be useful to distinguish more finely several different theses here which often seem to get conflated in the debate. To begin with, there is:

1. Strong Local Mechanism: The set of humanly provable mathematical truths is equivalent to the set of theorems of a certain explicitly specified formalized theory *F*: "this *F"*.

In other words, the mechanist is here supposed to explicitly present a particular formalized theory *F* which is contended to be equivalent with the human mind. However, it is clearly possible to advocate mechanism as a general thesis without such a specific claim:

2. Basic General Mechanism: The set of humanly provable mathematical truths is *equal* in effect to the set of theorems of *some* formalized theory.

Finally, we should also distinguish the following, apparently weaker, claim:

3. Weak General Mechanism: The set of humanly provable truths is *contained* in the set of theorems of *some* formalized theory.

There also appear to be several different anti-mechanist claims on offer.

4. Weak Anti-Mechanism: It follows from Gödel's incompleteness theorems that Strong Local Mechanism is false.

5. Basic Anti-Mechanism: It follows from Gödel's incompleteness theorems that Basic General Mechanism is false.

[5] Roughly, F_1 is interpretable in F_2 if the language of F_1 can be "translated" into the language of F_2 in such a way that F_2 proves the translation of every theorem of F_1. This notion of interpretability was first given an explicit definition by Tarski in (Tarski, Mostowski & Robinson 1953). It had been, however, already used in practice by logicians for some time.

6. Strong Anti-Mechanism: The human mind can surpass *any given* consistent formalized theory (which includes arithmetic) and prove ("see to be true") the Gödel sentence of it.[6]

It seems that Lucas, Penrose and their allies do not always sufficiently distinguish these different theses, but slide from one to another and back again without clearly noticing this. When Lucas, for example, declared that "given *any* machine which is consistent and capable of doing simple arithmetic, there is a formula it is incapable of producing as being true ... but which we can see to be true" (Lucas 1961; my emphasis), he apparently advocated Strong Anti-Mechanism.[7] However, when pressed, Lucas and others often retreat to Weak Anti-Mechanism, or perhaps to an even more specific view. That is, especially when anti-mechanists attempt to circumvent critique, the mechanist view they apparently focus on is even stronger and more specific than Strong Local Mechanism, namely:

7. Naïve Strong Local Mechanism:
(i) Strong Local Mechanism;
(ii) the human mind knows the equivalence of itself and the specific formalized theory *F* with *mathematical certainty* (i.e., the equivalence is itself absolutely provable);
(iii) the human mind knows with *mathematical certainty* that *F* is consistent.

We can grant Lucas and other anti-mechanists that Naïve Strong Local Mechanism collapses, in the light of the Gödelian facts, into inconsistency. This was already apparent for Gödel himself (see Gödel 1951) and has been repeatedly conceded. But this concession is a rather minute victory for anti-mechanism. For there are many ways to be a coherent mechanist without committing oneself to the Naïve Strong Local Mechanism.[8]
First, and most obviously, one might have general theoretical or empirical reasons for advocating Basic General Mechanism (or Weak General Mechanism), but not Strong Local Mechanism. But what is more, one might perhaps have *inductive empirical reasons* for believing that a particular formalized theory *F* corresponds to the human mind, but such reasons are, of course, short of mathematical certainty. On second thought, this seems a much more plausible alternative than the idea that it should be known with mathematical

[6] This is also the first disjunct of Gödel's famous, more cautious disjunctive thesis, now standardly called "Gödel's disjunction" (See Gödel 1951); the second disjunct says that there are mathematical problems which are absolutely undecidable for the human mind. Gödel suggested that their disjunction follows from the incompleteness results; but he never contended that the first disjunct in itself would follow.

Although our Strong Anti-Mechanism is not formulated directly as the opposite of Weak General Mechanism, it is natural to interpret the former as denying the latter.

[7] Note that Lucas here only requires that the machine *be* consistent – not that the mechanist, we or anyone *know* (with mathematical certainty) that it is consistent.

[8] Koellner (2016, 2018a), building on the earlier work of Reinhardt (1985a, b) and Carlson (2005), analyzes some such differences much more rigorously in the context of so-called epistemic arithmetic. He labels roughly the same view I have here called "Basic General Mechanism" as "weak mechanistic thesis"; Reinhardt (1985b) proved that it is consistent. The view that the former is itself knowable with mathematical certainty is called in this tradition the "strong mechanistic thesis" (this view does not occur separately in my listing above); Carlson (2005) showed that it is consistent. Finally, Koellner calls the view roughly corresponding to our Naïve Strong Local Mechanism "super strong mechanistic thesis"; it was proved inconsistent, in this context, by Reinhardt (1985a).

certainty. Finally, a mechanist might believe in the consistency of *F*, but on grounds that are weaker than absolute mathematical certainty, for example, broadly speaking *inductive*: *F* seems to avoid known paradoxes, no contradiction has so far been derived in it, it has some expected consequences etc.[9] (more of the latter below).

That is, even Weak Anti-Mechanism is as such false, and mere Strong Local Mechanism is not necessarily refuted by Gödel's theorems, unless it is complemented with the further conditions (ii) and (iii) from the definition of Naïve Strong Local Mechanism, and the latter is thus adopted. Hence, there is plenty of room for different mechanistic views which are not vulnerable to any Gödelian counterarguments; and if Weak Anti-Mechanism fails, Basic Anti-Mechanism and Strong Anti-Mechanism are on an even weaker footing.

Questions of Consistency

The standard objection[10] to the Gödelian anti-mechanist arguments builds on the fact that Gödel's first incompleteness theorem has in reality a conditional form, and the alleged truth of the Gödel sentence G_F for a formalized theory *F* depends on the assumption of the consistency of *F*. Therefore, in order to really know that G_F is true one must first know that *F* is consistent.[11] And that is not, in general, transparent.

Lucas and some of his devotees (but also some critics) seem to think that the gist of the objection is to raise doubts about the consistency of the human mind; but I think this is off the mark. The central notion here is *absolute provability* – what the human mind can prove with mathematical certainty. (In this paper, I use "absolutely provable" and "knowable with mathematical certainty" interchangeably.) Whatever the scope of such knowledge really is, this is a normative concept, and it is not terribly implausible to contend that it consists, by definition, of true sentences and is consequently a consistent whole. The real question is whether this totality of absolutely provable sentences is, by its very nature, such that it cannot, as a matter of absolute mathematical fact, coincide with or be contained in a set of theorems of some formalized theory.

In other words, the critical question is not whether I am consistent and/or whether I can know that I am consistent, but whether a given formalized theory is consistent and whether I can always know with mathematical certainty that it is. The challenge is especially flagrant for Strong Anti-Mechanism. Lucas explicitly contends that the human mind can surpass *any* consistent formalized theory and intuitively prove as true its Gödel sentence. However, that amounts to being able to prove intuitively and absolutely, with mathematical certainty, the consistency of any given formalized theory, if it is in fact consistent. And that is fantastically

[9] Gödel himself, in his Gibbs lecture (Gödel 1951), was already sensitive to these further conditions (i.e. (ii) and (iii)), when he qualified that the soundness (and, consequently, consistency) of the formalized theory should be known with "mathematical certitude", and reflected the possibility that the human might well know its equivalence with a formalized theory, but only with "empirical certainty".

[10] The objection goes back to Putnam (1960).

[11] And we know, from Gödel's second incompleteness theorem, that (under certain general conditions) the consistency of *F* cannot be proved inside *F*. In fact, it can be shown that the Gödel sentence G_F for *F* and the formalized consistency statement for *F* are materially equivalent inside *F* (and hence equally unprovable in *F*); see e.g. Raatikainen 2020.

optimistic indeed. Lucas and his followers greatly underestimate the difficulty of this task. Consistency is (in terms of computability theory) a Π_1^0-complete property. This means that being able to tell whether a given formalized theory is consistent or not would enable one to tell about *every* Π_1^0 sentence[12] whether it is true or false: one should have an "oracle"[13] for this class of sentences.[14] There is absolutely no reason to believe that the human mind has such miraculous powers. There are many open problems in mathematics which have this form (that is, Π_1^0). Even the best mathematicians have no clue how to know whether they are true or not; the same holds for the corresponding consistency questions.

Lucas (1961) writes, referring now to Gödel's second incompleteness theorem:

> All that Gödel has proved is that a mind cannot produce a formal proof of the consistency of a formal system inside the system itself: but there is no objection to going outside the system and no objection to producing informal arguments for the consistency either of a formal system or of something less formal and less systematized. Such informal arguments will not be able to be completely formalized: but then the whole tenor of Gödel's results is that we ought not to ask, and cannot obtain, complete formalization.

However, either such informal arguments of the human mind for the consistency of a formalized theory are less certain than absolute provability, which is (as we have noted above) perfectly compatible with mechanism. Or they have essentially the epistemological status of mathematical certainty, in which case Lucas's claim here amounts to the extremely strong claim that the human mind can access absolutely certain mathematical proofs which are in principle impossible to formalize – a strong claim badly in need of an argument for its support. It is something much stronger than what the Gödelian argument – even if it were successful – would provide. We may assume that in such proofs, pure logic, e.g. many-sorted first-order logic, is fixed, and everything else is given as non-logical axioms. Lucas's claim then implies that the human mind is able to use in its absolute proofs axioms which are somehow, in principle, impossible to formalize. The idea is baffling, and certainly Gödel's theorems entail no such thing.

Be that as it may, what has perhaps misled many here is that textbook presentations of Gödel's incompleteness theorems often take as their starting point some arithmetical theory which is both very familiar and relatively weak. For such a natural weak theory, it is plausible to say that we know its axioms to be true and consequently consistent, with mathematical certainty. But that just is not the case with an arbitrary formalized theory; our intuition

---

[12] Π_1^0 sentences are, roughly, the purely universal formulas; more exactly, formulas of the form $\forall x_1 \forall x_2 \ldots \forall x_n\, A$, where A does not contain any unbounded quantifiers (A may contain bounded universal quantifiers $\forall x < t$ and bounded existential quantifiers $\exists x < t$). Both the Gödel sentence and the arithmetized consistency statement (their standard formalizations) have this form. (Hodes 1998) is a helpful survey of such classifications of sentences and sets.

[13] An "oracle" is a heuristic idea, due to Turing, in computability theory. In the realm of undecidable problems, it is simply stipulated that an oracle can always immediately give the correct answer for some fixed class of questions.

[14] If a Π_1^0 sentence S is in fact false, it can always be proved to be false in any formalized theory which contains Robinson Arithmetic **Q**. Consequently, if a superbeing could decide the consistency question for all formalized theories, it could in particular decide whether the formal system **Q** + S is consistent or not. But that amounts to deciding whether S is true or false.

(whatever that may be) may well say nothing about their consistency. The point that I want to emphasize is that our (the human mind's) confidence concerning the consistency of formalized theories is a *matter of degree* and varies massively depending on the theory.

The Lucasian anti-mechanism apparently contends that the human mind can informally and absolutely *prove* the Gödel sentence of *any* given formalized theory *F* (and, equivalently, the consistency of *F*) in exactly the same sense and with the same degree of *mathematical certainty* that we can prove, say, 2 + 2 = 4, or the fundamental theorem of arithmetic (i.e. the unique-prime-factorization theorem). But when *F* is, for example, an unfamiliar and extremely strong theory, this is just not credible.

In the case of weak Robinson Arithmetic **Q**, we tend to be absolutely certain that it is consistent, and that is easy to prove with core mathematics. With the first-order Peano Arithmetic **PA**, which includes the induction scheme, we are perhaps still almost as confident about its consistency. But when we go beyond predicativity to the full second-order arithmetic **PA2**, we may have at least a lingering doubt whether it is consistent. Although many mathematicians and logicians are, in their everyday work, prepared to lean on Zermelo-Frankel set theory with the axiom of choice **ZFC**, there may also be reasonable doubts about its consistency.[15] And when one moves on to add to it stronger and stronger axioms of infinity – involving inaccessible, measurable, compact and supercompact and whatever huge cardinals etc. – our confidence concerning the consistency of the resulting theory decreases. The only evidence we have for their consistency may be that they *seem* to formalize a consistent notion, they *seem* to avoid known paradoxes, and that one has not, so far, derived a contradiction from them. With some complex unprecedented formalized theories, our intuition may well be totally helpless. It would be implausible to contend that the epistemological status of the consistency claims would always be on an equal footing and that of absolute mathematical certainty in all such very different cases. And exactly the same holds with the respective Gödel sentences. It is a matter of degree and varies enormously.[16]

I submit that it is quite plausible that there are consistent formalized theories so complex and powerful that they would be simply incomprehensible for the human mind, and the human mind would have in particular no clue as to whether they were consistent or not. Some such formalized theory may prove everything that the human mind could ever prove – and perhaps much more. Note that such a formalized theory might look very different from our familiar theories of arithmetic. It might just be that our theories of arithmetic are relatively interpretable (see above) in such a theory – we might not even be able to see that this is the case[17] – and as such be able to prove every arithmetical truth the human mind could ever even in principle prove. Gödel's results are perfectly compatible with such a state of affairs.

---

[15] Obviously, these are just possible examples, and in real life, the attitudes of different mathematicians and logicians vary.

[16] Cf. Davis 1990a; Raatikainen 2005.

[17] Though some familiar interpretations (of a theory in another theory) are quite elementary, the general relation of relative interpretability is in fact highly undecidable: in logicians' terms, it is Σ_3^0 (Shavrukov 1997); it is thus not decidable even in the limit; and there are cases whose verification is, by all reason, beyond the capacities of the human mind.

Concluding Remarks

I think it is quite clear that the actual operation of the human mind, even in the realm of pure mathematics, differs in practice in several ways from a deterministic Turing machine (corresponding via a fixed coding to a formalized theory) which just mechanistically derives and enumerates theorems of some fixed axiom system in some systematic order.

Often, there is first a conjecture formulated in whatever creative way, and then varying attempts to prove it; with luck, ingenuity and hard work and after some dead ends, a proof may at some point be found. Sometimes conceptual revolutions take place in mathematics, as when mathematics moved from more computational and discrete approaches to analysis, with the notions of continuity and limit etc., and eventually to infinitary set theory. Sometimes mere inductive reasoning is, *faute de mieux*, used in support of a hypothesis, as Putnam (1975), for example, has pointed out. New axioms are sometimes tentatively accepted, not because they are seen to be true with absolute certainty, but only because they have some expected and desirable consequences, as Maddy (1988), among others, has emphasized. And so on. However, the claim at issue here has been whether Gödel's incompleteness results demonstrate that the human mind can surpass any given formalized theory; and none of the above observations make it any more the case.

We have noted that the notion of *absolute provability* is at the core of the debate. However, skepticism concerning this concept is emerging. I have emphasized above (see also Raatikainen 2005) that certainty in mathematics is a matter of degree and varies tremendously even among Π_1^0 sentences. But this implies that absolute provability and mathematical certainty do not have the sort of sharp on/off-boundaries that the Gödelian argument for anti-mechanism presupposes they have. Recently, several philosophers and logicians have expressed, in different but complementary ways, doubts about the very clarity of the concept of absolute provability in this context (see Koellner 2016, 2018b; Shapiro 2016, Williamson 2016). Upon closer scrutiny, it is suspect whether this notion is at all sufficiently well-defined. But if that is the case, so much the worse for the Gödelian anti-mechanist arguments.

Even if mechanism may suggest a somewhat distorted and misleading picture of the human mind in its mathematical mode, there is, nevertheless, some point in making an effort to criticize the popular Gödelian arguments against mechanism: they in turn suggest a highly unrealistic picture of both the powers of the human mind in mathematics and the powers of mathematical methods in establishing ambitious philosophical conclusions. Such an unfounded mystification of the human mind is certainly worth condemning. Exciting as Gödel's results are, they simply cannot do all the philosophical work they are often assigned to.

References

Boolos, George (1968). "Review of 'Minds, Machines and Gödel', by J.R. Lucas, and 'God, the Devil, and Gödel', by P. Benacerraf", *Journal of Symbolic Logic* 33, 613–15.

Carlson, Timothy J. (2005). "Knowledge, Machines, and Reinhardt's Strong Mechanistic Thesis", *Annals of Pure and Applied Logic* 105, 51–81.

Davis, Martin (1990). "Is Mathematical Insight Algorithmic?", *Behavioral and Brain Sciences* 13, 659–660.

Davis, Martin (1993). "How Subtle is Gödel's Theorem? More on Roger Penrose", *Behavioral and Brain Sciences* 16, 611–612.

Feferman, Solomon (1995). "Penrose's Gödelian argument: A Review of *Shadows of Mind*, by Roger Penrose," Psyche, 2 (7).

Feferman, Solomon (2009). "Gödel, Nagel, Minds, and Machines", J*ournal of Philosophy* 106 (4): 201–219.

Feferman, Solomon (2011). "Gödel's Incompleteness Theorems, Free Will and Mathematical Thought", in Richard Swinburne (ed.), *Free Will and Modern Science*. OUP/British Academy.

Franzén, Torkel (2005). *Gödel's Theorem: An Incomplete Guide to its Use and Abuse*, Wellesley: A.K. Peters.

Gaifman, Haim (2000). "What Gödel's Incompleteness Result Does and Does not Show", *The Journal of Philosophy* 97, 462–70.

Gödel, Kurt (1931). "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I", *Monatshefte für Mathematik und Physik* 38, 173–98; translated in Gödel 1986, 144–195.

Gödel, Kurt (1951) "Some Basic Theorems on the Foundations of Mathematics and their Implications" (Gibbs Lecture). In Gödel 1995, 304–323.

Gödel, Kurt (1986). *Collected Works I. Publications 1929–1936.* ed. S. Feferman *et al.*, Oxford University Press, Oxford.

Gödel, Kurt (1990). *Collected Works II. Publications 1938–1974.* ed. S. Feferman *et al.*, Oxford University Press, Oxford.

Hodes, Harold (1998). "Recursion-Theoretic Hierarchies", *Routledge Encyclopedia of Philosophy*, Taylor and Francis, https://www.rep.routledge.com/articles/thematic/recursion-theoretic-hierarchies

Horsten, Leon & Welch, Philip (eds.) (2016). *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*. Oxford: Oxford University Press.

Koellner, Peter (2016). "Gödel's Disjunction", in Horsten & Welch 2016, 148–188.

Koellner, Peter (2018b). "On the Question of Whether the Mind Can Be Mechanized, II: Penrose's New Argument", *Journal of Philosophy* 115, 453–484.

Koellner, Peter (2018a). "On the Question of Whether the Mind Can Be Mechanized, I: From Gödel to Penrose", *Journal of Philosophy* 115, 337–360.

Lucas, J. R. (1961). "Minds, Machines, and Gödel", *Philosophy* 36, 112–137.

Lucas, J. R. (1996). "Minds, Machines, and Gödel: A Retrospect", in P.J.R. Millican and A. Clark (eds.) *Machines and Thought. The Legacy of Alan Turing*, Vol. 1, Oxford University Press, Oxford, 103–124.

Maddy, Penelope (1988). "Believing the Axioms I, II", *Journal of Symbolic Logic* 53, 481–511, 736–64.

Megill, Jason (2012). "The Lucas-Penrose Argument about Gödel's Theorem", *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002. https://www.iep.utm.edu/lp-argue/

Nagel, Ernest and James R. Newman (1958). *Gödel's Proof*, New York University Press, New York.

Odifreddi, Piergiorgio (1989). *Classical Recursion Theory*, Amsterdam: North-Holland.

Penrose, Roger (1989). *The Emperors New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford University Press, New York.

Penrose, Roger (1994). *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford University Press, New York.

Penrose, Roger (1995). "Beyond the Doubting of a Shadow: A reply to Commentaries of *Shadows of the Mind*", *Psyche* Vol 2.

Penrose, Roger (1997). "On Understanding Understanding", *International Studies in the Philosophy of Science* 11, 7–20.

Putnam, Hilary (1960). "Review of *Gödel's Proof* by Ernest Nagel & James R. Newman", *Philosophy of Science*, Vol. 27, 205–207.

Putnam, Hilary (1975). "What is Mathematical Truth?", *Historia Mathematica* 2, 529–545. Reprinted in H. Putnam: *Mathematics, Matter and Method. Philosophical Papers Vol 1*. Cambridge University Press, Cambridge, 1975, 60–78.

Raatikainen, Panu (2005). "On the Philosophical Relevance of Gödel's Incompleteness Theorems", *Revue Internationale de Philosophie* 59, 513–534.

Raatikainen, Panu (2020). "Gödel's Incompleteness Theorems", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/sum2020/entries/goedel-incompleteness/>.

Reinhardt, William N. (1985a). "Absolute Versions of Incompleteness Theorems", *Nous* 19, 317--346.

Reinhardt, William N. (1985b). "The Consistency of a Variant of Church's Thesis with an Axiomatic Theory of an Epistemic Notion", in *Special Volume for the Proceedings of the 5th Latin American Symposium on Mathematical Logic, 1981, volume 19 of Revista Colombiana de Matem'aticas*, 177–200.

Shavrukov, V. Yu. (1997). "Interpreting reflexive theories in finitely many axioms", *Fundamenta Mathematicae* 152, 99–116.

Shapiro, Stewart (1998). "Incompleteness, Mechanism, and Optimism", *Bulletin of Symbolic Logic* 4, 273–302.

Shapiro, Stewart (2016). "Idealization, Mechanism, and Knowability", in Horsten & Welch 2016, 189–207.

Tarski, A., Mostowski, A., and Robinson, R.M. (1953). *Undecidable Theories*, Amsterdam: North-Holland.

Williamson, Timothy (2016). "Absolute Provability and Safe Knowledge of Axioms", in Horsten & Welch 2016, 243–253.