

Absolute Provability and Safe Knowledge of Axioms

(to appear in Leon Horsten and Philip Welch, eds., *The Limits of Mathematical Knowledge*, Oxford University Press)

Timothy Williamson

I

When mathematicians discuss proofs, they rarely have a particular formal system in mind. Indeed, they are typically not thinking of formal systems at all, although they might accept the suggestion that a genuine proof can in principle be reconstructed in an appropriate formal system. The picture is more like this. At any given time, the mathematical community has a body of knowledge, including both theorems and methods of proof. Mathematicians expand mathematical knowledge by recursively applying it to itself, adding new theorems and sometimes new derived methods of proof. Of course, present mathematical knowledge itself grew out of a smaller body of past mathematical knowledge by the same process. Since present mathematical knowledge is presumably finite, if one traces the process back far enough, one eventually reaches ‘first principles’ of some sort that did *not* become mathematical knowledge in that way. Some principles

of logic and axioms of set theory are good candidates for such first principles. They became human mathematical knowledge within the last few thousand years, in some cases within the last century or two, but in some other way. Most mathematicians are hazy about what such other ways might be. Let us apply the term ‘normal mathematical process’ to all those ways in which our mathematical knowledge can grow. Normal mathematical processes include both the recursive self-application of pre-existing mathematical knowledge and the means, whatever they were, by which first principles of logic and mathematics originally became mathematical knowledge. By contrast, appeals to mere inductive support or alleged divine testimony presumably do not count as normal mathematical processes.

A mathematical hypothesis is *absolutely provable* if and only if it can in principle be known by a normal mathematical process. When mathematicians discuss provability, they typically have in mind something like absolute provability. No assumption is made either way as to whether a single formal system generates all and only the absolutely provable hypotheses.

Since, necessarily, only truths are known, absolute provability entails possible truth. Therefore, since mathematical hypotheses are entirely non-contingent and so true if possibly true, for them absolute provability entails actual truth (where ‘necessarily’ and ‘possibly’ express metaphysical rather than epistemic modalities).

A mathematical hypothesis is *absolutely decidable* if and only if either it or its negation is absolutely provable; otherwise it is *absolutely undecidable*. Suppose that every mathematical hypothesis is either true or false; in the latter case, it has a true negation. Then every mathematical hypothesis is absolutely decidable if and only if every

true mathematical hypothesis is absolutely provable: in other words, truth coincides with absolute provability for mathematical hypotheses. If every mathematical hypothesis is absolutely decidable, then for Gödelian reasons no single formal system generates all and only the absolutely provable hypotheses.

Are there absolutely undecidable mathematical hypotheses? Before attempting to answer that question, we must further clarify its terms. The next section contributes to that process.

II

The first clarification is mainly a matter of book-keeping. On the natural reading, a mathematical hypothesis is a *proposition*. A proposition is not a sentence, although it can be expressed by a sentence, indeed, by different sentences in different languages.

Goldbach's Conjecture is the same proposition whether expressed in English, German, or an interpreted formal language. By contrast, the theorems of a formal system are standardly equated with the formulas themselves, not with the propositions they may express.

Unfortunately, some prominent theories of propositions are so coarse-grained as to make the propositions literally expressed by mathematical formulas useless for epistemological purposes. More specifically, if one follows Robert Stalnaker in treating a proposition as the set of (metaphysically) possible worlds at which it is true, then all true mathematical formulas literally express the same proposition, the set of all possible

worlds, since all true mathematical formulas literally express necessary truths. It is therefore trivial that if one true mathematical proposition is absolutely provable, they all are. Indeed, if you already know one true mathematical proposition (that $2 + 2 = 4$, for example), you thereby already know them all. Stalnaker suggests that what mathematicians really learn are in effect new contingent truths about which mathematical formulas we use to express the one necessary truth, but his view faces grave internal problems, and the conception of the content of mathematical knowledge as contingent and metalinguistic is in any case grossly implausible.¹ A subtler alternative within the framework of coarse-grained propositions is to interpret mathematicians as coming to know the same old necessary truth under the guise or mode of presentation of a new formula. However, many philosophers of language react to the problem by appealing to finer-grained theories of propositions, although they face their own difficulties.² For present purposes, we may put the problem to one side, by treating the bearers of absolute provability as interpreted formulas rather than propositions, despite the unnaturalness of making formulas the objects of knowledge-that.

A more urgent need for clarification concerns the scope of the phrase ‘normal mathematical process’. If our interest is in making mathematical progress over the next few decades, for example in deciding the Continuum Hypothesis (CH) and other issues undecidable within our currently accepted formal systems, then what matter are the mathematical processes feasible for humans physically and psychologically not too different from ourselves, in social and cultural circumstances not too different from our own. For practical purposes, it is presumably pointless to speculate about proofs or disproofs of CH graspable by creatures with brains much larger than ours (but still finite),

unless we treat computers or humans-plus-computers as such creatures. For more theoretical purposes, however, it is rather arbitrary to postulate a specific finite upper bound on the intellectual capacity of the mathematicians with respect to whom ‘normal mathematical process’ is defined. Any such bound is likely to be incompatible with the closure of normal mathematical processes under natural operations. For example, we cannot assume that provability within such a bound is closed under modus ponens, since even if a proof of $A \rightarrow B$ and a proof of A both fall below a finite complexity threshold, the result of combining them into a proof of B may fail to do so. Such an accidentally contoured standard of provability is unlikely to be theoretically very rewarding. Consequently, the phrase ‘normal mathematical process’ will here be understood without any specific finite bound or restriction to the physical, psychological, social, and cultural limitations characteristic of humans, now and perhaps even in the distant future. However, we will understand ‘normal mathematical process’ as restricted to the capacities of mathematicians whose operations are finite in some appropriate sense. Checking Goldbach’s Conjecture for every natural number one after another, the check on $n + 1$ of course taking half the time of the check on n , does not count.

III

One normal mathematical process, even if a comparatively uncommon one, is adopting a new axiom. If set theorists finally resolve CH, that is how they will do it. Of course, just arbitrarily assigning some formula the status of an axiom does not count as a normal

mathematical process, because doing so fails to make the formula part of mathematical *knowledge*. In particular, we cannot resolve CH simply by tossing a coin and adding CH as an axiom to ZFC if it comes up heads, \sim CH if it comes up tails. We want to *know* whether CH holds, not merely to have a true or false belief one way or the other (even if we could get ourselves to believe the new axiom). Thus the question arises: when does acceptance of an axiom constitute mathematical knowledge?

In answering the question, we must not allow ourselves to be distracted by the use of the word ‘axiom’ to describe clauses in the definitions of various classes of mathematical structure, such as groups, rings, fields, and topological spaces. Axioms in that sense do not even express independent propositions, and the question of knowledge does not arise, although we can of course ask whether someone is familiar with the definition. Rather, in the present and more traditional sense, an axiom is a principle relied on without further proof. The usual phrase is just ‘without proof’, but a standard definition of a proof makes the sequence whose only member is an axiom (in the relevant sense) constitute a proof of that very axiom, and we may accept that more elegant way of thinking. The practice of proof in mathematics requires some axioms in that sense, some first principles, even if which principles have that status is sometimes in dispute.

If some form of logicism is correct, on a liberal conception of logic, then mathematical proof can make do with purely logical first principles. Even if mathematics requires some non-logical first principles, it may be possible to resolve outstanding mathematical problems such as CH by adding new first principles all of which are purely logical: for instance, principles of second-order logic (which is incomplete). For present purposes, we can simply treat logical reasoning as a normal mathematical process, and be

neutral as to whether the first principles at issue are purely logical or distinctively mathematical.

To start a line of thought, let A be a true interpreted mathematical formula that cannot be proved with just the resources of current human mathematics. When presented with A , human mathematicians are simply agnostic.

We may assume here that mathematical truths are not contingent: neither the existence nor the non-existence of absolutely unprovable mathematical truths puts that into question. Thus A is a (metaphysically) necessary truth: in the relevant respect, things could not have been otherwise. Obviously, that does not mean that the formula A , if individuated typographically, could not have had a quite different meaning and expressed a falsehood. Such counterfactual interpretations are always possible, but irrelevant. For example, ' $0 = 0$ ' would have expressed a falsehood if we had used '=' to mean *is less than*. What matters is that A as *actually* interpreted expresses a necessary truth. If we introduce the operator \Box to mean 'It is necessary that', we can put the point by saying that $\Box A$ is true.

For epistemological purposes, the upshot is that error is impossible in believing A when it expresses what it actually expresses.³ But that does not mean that whoever believes A on its actual interpretation thereby knows A . For instance, if I am gullible enough to believe whatever my guru tells me, and he decides on the toss of a coin to tell me A rather than $\sim A$, I do not thereby know A , let alone have specifically *mathematical* knowledge of A . I could far too easily have come to believe the falsehood $\sim A$ in a relevantly similar way.

However, there could in principle be mathematicians, perhaps non-human ones, who believe A and could *not* easily have come to believe $\sim A$ or any other falsehood in a relevantly similar way. As a by-product of the evolutionary history of their kind, their brains have come to be wired so as strongly to predispose them to accept as obvious any formula which they interpret as we actually interpret A , when the formula is presented to them. As is sometimes said, they find such formulas ‘primitively compelling’.⁴ For simplicity, we may suppose that the specific feature of their brains that causes them to find a formula with A ’s meaning primitively compelling does not cause them to find any other formulas primitively compelling. Their brains could not at all easily have been wired differently enough for them to lack that feature. These creatures could not at all easily have had a false belief in A on its actual interpretation, nor have we any reason to suppose that they could easily have had a false belief in A on some other interpretation. Do the creatures know A , once it is presented to them on its actual interpretation?

One might be tempted to deny that the creatures know A , on the grounds that their belief in A is not appropriately connected with its truth. But what sort of ‘appropriate connection’ is being demanded here? Doubtless, if the creatures are asked ‘Why does A hold?’, they can only stutter and say ‘It’s just obvious’, or perhaps glibly produce some rigmarole that begs the question in favour of A . But isn’t that the very reaction that normal people give when asked why their first principles hold, principles that they take themselves to know?

Let us elaborate the story. We may suppose that the creatures develop a form of mathematics that includes all the methods and results of current human mathematics, together with A , suitably interpreted, as an additional axiom. Thus their mathematics is

strictly somewhat more powerful than ours. They use A to settle various previously open questions. Since both A and all of their other first principles are true, no inconsistency ever results. Indeed, A will have various consequences that the creatures can independently test on the basis of the A -free part of their mathematics; of course, those consequences always pass the test. The creatures also have a vague sense that A coheres with their other first principles into a unified picture of mathematical reality, although they are unable to substantiate that claim in any rigorous way. Do they lack anything with respect to A that we have with respect to our first principles of logic and mathematics?

Some philosophers think that the key to the epistemology of first principles of logic is that they are *analytic*, in the sense that a disposition to assent to them is essential to linguistic or conceptual competence with the logical constants that figure in them.⁵ Clearly, A is not analytic in that sense, since by hypothesis we interpret A as the creatures do but are not disposed to assent to it. But it is very doubtful that *any* principles of logic are analytic in that sense. Someone may acquire the relevant logical constants in the normal way, and have normal linguistic or conceptual competence with them, but come to reject the principle at issue on sophisticated (albeit mistaken) theoretical grounds, and even find the original disposition to assent to the principle morphing into a different principle to assent to some more qualified principle, while still retaining normal competence with those logical constants, and using them within the bounds of everyday normality.⁶ The point is even clearer for mathematical expressions. A well-trained set theorist may come to challenge one of the axioms of ZFC on theoretical grounds, and cease to find it primitively compelling, without incurring any credible charge of linguistic or conceptual incompetence with the word 'set' or the ' \in ' symbol for set membership.

The case for the analyticity of first principles of mathematics in the relevant sense does not stand up. Thus *A* should not be disqualified from the status of a mathematical axiom on grounds of its non-analyticity. Nor is there any good reason to deny that the creatures use the constituent expressions of *A* with the same meanings as we do. Their mathematical practice subsumes ours: that they know more than we know does not imply a difference in reference. For example, if the relevant mathematical language is that of arithmetic, it would be bizarre to suggest that ‘+’ refers to a different operation from the one it refers to in our mouths.

A more moderate constraint on mathematical axioms is that they should be known *a priori*. They are supposed to be known independently of experience, in some appropriate sense. The nature and significance of the distinction between *a priori* and *a posteriori* knowledge are much in dispute.⁷ For present purposes, however, we may simply note that the creatures’ knowledge of *A* no more depends on experience than does human knowledge of our mathematical axioms. Perhaps the experience of the creatures’ distant ancestors played some role in the evolutionary history of the brain structures underlying the creatures’ assent to *A*, but then the experience of our distant ancestors surely played some role in the evolutionary history of the brain structures underlying our assent to our mathematical axioms. The human capacity for mathematics did not come about by magic or divine inspiration: it is rooted in our capacity to recognize the shapes and compare the sizes of external objects, to divide and combine those objects in different ways, to sort them according to different principles, to permute them and count them, and so on. If our knowledge of our mathematical axioms is *a priori*, then so too is the creatures’ knowledge of *A*.

At least provisionally, we should admit that the creatures' knowledge of A is no worse than our knowledge of our axioms. In current epistemological terms, their knowledge of A meets the condition of *safety*: they could not easily have been wrong in a relevantly similar case. Here the relevantly similar cases include cases in which the creatures are presented with sentences that are similar to, but still discriminably different from, A , and express different and false propositions; by hypothesis, the creatures refuse to accept such other sentences, although they may also refuse to accept their negations.⁸ Thus A is fit to be a mathematical axiom for the creatures, if not for human mathematicians. Therefore A is absolutely provable, because the creatures can prove it in one line. It would be pointless to disqualify the proof for its triviality, for the story could just as well have had another true formula B instead of A playing the axiomatic role in the creatures' mathematics, where both the creatures and we can prove $B \rightarrow A$, but only with some difficulty.

The claim is not just that A *would* be absolutely provable *if* there were such creatures. The point is the stronger one that A *is* absolutely provable because there *could* in principle be such creatures. Absolute provability is not supposed to be contingent on what creatures happen to be around: it is not subject to the accidents of evolutionary history.

Since the foregoing argument made no special assumptions about the formula A , it generalizes: every true formula of mathematics is absolutely provable, and every false formula is absolutely refutable. Thus every bivalent mathematical formula is absolutely decidable. If every mathematical formula is either true or false, then every mathematical formula is absolutely decidable. At least for the language of first-order arithmetic, it is

overwhelmingly plausible that every formula *is* either true or false, so in particular every formula of the language of first-order arithmetic is absolutely decidable.⁹

IV

The absolute provability of all arithmetical truths in the sense above obviously does not imply that human or humanoid minds can somehow do more than machines. It means that for every arithmetical truth A , it is possible for a finitely minded creature somewhat like us to prove A (in the absolute sense). But nothing has been said to exclude the hypothesis that for every arithmetical truth A , some implementation of a Turing machine can prove A (in the absolute sense). That is not just the trivial claim that for every arithmetical truth A , some implementation of a Turing machine can print out A . Rather, the implementation of the Turing machine is required to come to *know* A by a normal mathematical process, and so to be minded. Presumably, on a suitable understanding of the relevant terms, no possible implementation of a Turing machine can prove *all* arithmetical truths. But nothing has been said to support the claim that some possible finitely minded creature somewhat like us *can* prove all arithmetical truths. More generally, no relevant asymmetry has been proposed between finitely minded creatures somewhat like us and implementations of Turing machines.

The possibilities under consideration do not form a linear structure, such as an idealized time sequence in which knowledge grows cumulatively and every truth of the mathematical language is sooner or later known. Rather, they form a branching tree

structure in which what is known at one point may neither include nor be included by what is known at another. If the points are time-indexed, it is not even required that along some or every possible history (an infinite path through the tree, taking in every time once) every truth is sooner or later known; nor is it required that knowledge grows cumulatively. We simply identify what is absolutely provable with what is known somewhere or other in the whole branching structure. The truths known at any given point are recursively axiomatizable; the truths known somewhere or other in the structure are not. The coincidence of absolute provability and truth for the mathematical language is even consistent with the recursive axiomatizability of the totality of truths ever known along any given possible history, so that not every mathematical truth ever gets known along any given possible history. For example, in some models every possible history reaches a point after which knowledge stops growing, but every mathematical truth is known in some possible history or other. However, no special reason has emerged for doubting the possibility of a history in which every mathematical truth is sooner or later known.

Someone might nevertheless speculate that some possible finitely minded creature might be able to prove all arithmetical truths, by having a brain structure that somehow encoded a non-recursive pattern recognition capacity for true formulas of arithmetic in an understood notation. But such a scenario goes far beyond the rather mundane extensions of human cognitive capacity on which the argument above relies. For present purposes, we need not indulge in such wild speculations.

We are in no danger of anti-mechanist conclusions even if we grant, for the sake of argument, that the possible finitely minded creatures somewhat like us are indeed

possible *humans*, members of our species, perhaps at some later stage of evolution. Then we have that for every arithmetical truth A , it is possible for there to be a human who in the factive absolute sense can prove A . But we still do not have that it is possible for there to be a human who, for every arithmetical truth A , can in that sense prove A . There is still no asymmetry between humans and implementations of Turing machines.

In this respect, Gödel's generic talk of 'the human mind' in his Gibbs lecture is dangerously misleading: '*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*'.¹⁰ For if the mathematical powers of 'the human mind' are understood as comprising all those mathematical powers it is possible for a human to have, then by parity of definition the mathematical powers of 'the finite machine' should comprise all those mathematical powers it is possible for a finite machine to have. Then, on our envisaged scenario, although Gödel's first disjunct is true, because within the realm of pure mathematics 'the human mind' infinitely surpasses the powers of any finite machine, it is equally true that within the realm of pure mathematics 'the finite machine' surpasses the powers of any finite machine. For just as any diophantine problem can in principle be solved by some human, although no human can solve them all, so any diophantine problem can in principle be solved by some finite machine, although no finite machine can solve them all. On the other hand, if the mathematical powers of 'the human mind' are understood as restricted to those mathematical powers that any first-rate human mathematician would have if granted infinite supplies of pencils, paper, and life without mental decay, then although Gödel's first disjunct is false on the envisaged scenario, the second disjunct is true only with

respect to those limited powers, in a sense compatible with every diophantine problem's being solvable in principle by some possible human or other, typically with greater mathematical powers than those of 'the human mind' as just redefined. Gödel's use of the generic definite article obscures crucial quantificational structure.

Talk of the powers of 'the human mind' may work better within a conception on which all normal humans have the same intellectual competence, all differences coming from accidental limitations on performance. The argument in section III did not suggest that the envisaged knowledge of axioms would be attainable within the limits of such normal human competence. But it is hard to see why evolutionary processes should be *in principle* incapable of giving our distant descendants non-accidentally greater intellectual powers than we have, just as we presumably have non-accidentally greater intellectual powers than our distant ancestors had. Whether those descendants count as humans, members of our species, is a zoological question of only moderate epistemological interest.¹¹

V

Some readers may find the argument in section III for the absolute provability of mathematical truths quite dissatisfying. It lacks both the mathematical difficulty and the epistemological depth one might expect an argument for such a conclusion to need. What raises those expectations? At least two factors are relevant.

First, the argument does not help *us* solve recalcitrant mathematical problems. We may be incapable of mathematical knowledge of the new axioms. We may be capable of knowing them only on the testimony of others, perhaps creatures of a different kind, and in any case we are virtually certain not to meet those creatures. To help *us* attain mathematical knowledge of new axioms, an argument surely would need to engage in detail with issues of great mathematical difficulty. But once we separate the practical question ‘How can we know the new axioms?’ from the theoretical question ‘Can the new axioms be known?’, we should not expect the latter to inherit all the difficulties of the former.

Second, the account in section III of possible knowledge of the new axioms does not fit the traditional picture of the epistemology of mathematical proof. Gödel emphasizes that he is talking about knowledge by purely mathematical means, not just knowledge by any means of purely mathematical propositions. He contrasts ‘mathematical certainty’ with ‘empirical certainty’.¹² In section III, the new axioms were envisaged as known by an ordinary standard of safety, not by some extraordinary standard of epistemic certainty unattainable elsewhere. In Gödel’s own case, it is anyway not clear how much weight he put on ‘mathematical certainty’, since he did not require mathematical intuition to be fallible.

For anyone who objects to the argument above that it does not involve mathematical certainty of the new axiom, the challenge is to explain the nature of this ‘mathematical certainty’ we are supposed actually to have for the current axioms but lack for the new one in the hypothetical scenario. It is not the necessity of the axioms, because that is the same in the two cases. Nor is it their subjective certainty, the doxastic state of

the agent reported by ‘I am certain that P’ rather than the epistemic status of the proposition reported by ‘It is certain that P’, for subjective certainty, unlike objective certainty, is not even factive: whereas if it is certain that P it follows that P, if someone is certain that P it does *not* follow that P. In any case, the subjective certainty may also be the same in the two cases; the new axiom is just as primitively compelling for the hypothetical agent as the current axioms actually are for us. Nor does the crucial difference consist in whether the knowledge is *a priori*, since we saw it to be so in both cases by the usual standard. Epistemological analyticity is also not the difference, since we saw it to be absent in both cases. Although one may feel tempted to start using phrases like ‘mathematical intuition’, ‘rational insight’, or ‘self-evidence’, without further explanation they merely obscure the issue, because it is so unclear that they add anything of epistemological significance to the distinctions already made. Even if they imply some distinctive extra phenomenology, we may legitimately suppose it to be present in the possible creatures who know the new axiom.

I am content to leave the matter with this challenge: if mathematical knowledge of axioms takes more than section III assumed, exactly *what* more does it take?

Acknowledgement. An early version of this material was presented at the 2012 Workshop on the Limits and Scope of Mathematical Knowledge at the University of Bristol. Later versions were presented in talks at Tel Aviv University, the University of Michigan at Ann Arbor, and the University of Oxford. I thank the audiences for helpful discussion, and Leon Horsten for useful written comments.

timothy.williamson@philosophy.ox.ac.uk

New College

Oxford OX1 3BN

U.K.

Notes

1 See Stalnaker 1999. One problem for his view is that for any formula A in standard mathematical notation, the biconditional $\text{True}(\ulcorner A \urcorner) \leftrightarrow A$ will follow logically from known axioms of a standard compositional theory of truth for the mathematical language. Since on Stalnaker's view our knowledge is closed under such logical consequence, it implies that we already know the biconditional, so the metalinguistic claim is equivalent for us to the original mathematical claim. Thus semantic ascent to the metalinguistic level only postpones the problem. I believe that Saul Kripke made a similar objection to Stalnaker's view.

2 For example, Russellian theories of structured propositions have trouble with Russellian paradoxes. Let $\langle O, p \rangle$ be the structured proposition that results from applying the propositional operator O to the structured proposition p . Let N be propositional negation. Define a propositional operator R such that $\langle R, \langle O, p \rangle \rangle$ is equivalent to $\langle N, \langle O, \langle O, p \rangle \rangle \rangle$ for any such operator O and proposition p (let $\langle R, q \rangle$ be equivalent to q if q is not of the form $\langle O, p \rangle$). But then $\langle R, \langle R, p \rangle \rangle$ is equivalent to $\langle N, \langle R, \langle R, p \rangle \rangle \rangle$, its own negation. Although measures can be taken to block such paradoxes, they significantly complicate the theory of propositions. Since banning all iterations of propositional operators would be much too restrictive, one needs paradox-free ways of defining the effect of an operator O on a structured proposition p of which O may itself be a constituent. The problem does not arise for more coarse-grained theories,

since when an operator O is applied to an unstructured proposition p to yield an unstructured proposition $O(p)$, there is no general way of uniquely recovering O and p from $O(p)$, so the analogue of R is manifestly ill-defined.

3 Of course, an indexical sentence such as ‘I am hungry’ can express a different proposition with a different truth-value in a different context without changing its linguistic meaning (its character, in the sense of Kaplan 1989). But mathematical formulas do not normally exhibit such indexicality. Arguably, many sentences containing a vague word such as ‘heap’ could easily have had slightly different linguistic meanings, and expressed different propositions with different truth-values, if the vague word had been used slightly differently (Williamson 1994). But that phenomenon too is not relevant to normal mathematical ignorance, for example when the sentence at issue is in the language of first-order arithmetic.

4 See Peacocke 1992, although in the present paper the phrase is used without commitment to Peacocke’s theory of concepts.

5 See Boghossian 2003.

6 See Williamson 2007 and, for a recent exchange, Boghossian 2012 and Williamson 2012.

7 For an argument that the distinction between *a priori* and *a posteriori* does not correspond to a deep epistemological difference see Williamson 2013.

8 See Williamson 2000. Williamson 2009 provides apparatus for extending the safety conception from sentences to arguments.

9 The thought experiment sketched above is similar in spirit to the thought experiment about the mathoids developed for a different but related purpose in Berry 2013. The two papers were written independently of each other; their conclusions are different, but compatible. Berry considers in detail various attempts to find an epistemological asymmetry between the mathoids and us, and argues convincingly that they all fail. Analogues of her considerations are relevant to the present case too.

10 Gödel 1951, p. 310. For good discussion see Shapiro 1998 and Feferman 2006.

11 Perhaps Gödel individuated human minds more flexibly than he did machines, treating the former as able to preserve their identity over time through changes that would not preserve the identity of a machine. But such an asymmetry is not motivated by, and does not depend on, his incompleteness theorems.

12 Gödel 1951, p. 309.

References

- Almog, Joseph, Perry, John, and Wettstein, Howard (eds.) 1989: *Themes from Kaplan*.
Oxford: Oxford University Press.
- Berry, Sharon 2013: 'Default reasonableness and the mathoids', *Synthese*, 190, pp. 3695-3713.
- Boghossian, Paul 2003: 'Blind reasoning', *Aristotelian Society* sup. 66, pp. 225-248.
- Boghossian, Paul 2012: 'Inferentialism and the epistemology of logic: reflections on Casalegno and Williamson', *Dialectica* 66, pp. 221-236.
- Casullo, Albert, and Thurow, Joshua (eds.), 2013: *The A Priori in Philosophy*. Oxford: Oxford University Press.
- Feferman, Solomon 2006: 'Are there absolutely unsolvable problems? Gödel's dichotomy', *Philosophia Mathematica* 14, pp. 134-152.
- Gödel, Kurt 1951: 'Some basic theorems on the foundations of mathematics and their implications'. In Gödel 1995, pp. 304-323.
- Gödel, Kurt 1995: *Collected Works, Vol. III. Unpublished Essays and Lectures*. S. Feferman, et al. (eds.). New York: Oxford University Press.
- Kaplan, David 1989: 'Demonstratives'. In Almog, Perry, and Wettstein 1989, pp. 481-563.
- Peacocke, Christopher 1992: *A Study of Concepts*. Cambridge, Mass.: MIT Press.
- Shapiro, Stewart 1998: 'Incompleteness, mechanism, and optimism', *The Bulletin of Symbolic Logic* 4, pp. 273-302.
- Stalnaker, Robert 1999: *Context and Content*. Oxford: Oxford University Press.

- Williamson, Timothy 1994: *Vagueness*. London: Routledge.
- Williamson, Timothy 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.
- Williamson, Timothy 2007: *The Philosophy of Philosophy*. Oxford: Blackwell.
- Williamson, Timothy 2009: 'Probability and danger', *Amherst Lectures in Philosophy* 4, pp. 1-35. <http://www.amherstlecture.org/williamson2009/>
- Williamson, Timothy 2012: 'Boghossian and Casalegno on understanding and inference', *Dialectica* 66, pp. 237-247.
- Williamson, Timothy 2013: 'How deep is the distinction between a priori and a posteriori knowledge?' In Casullo and Thurow 2013, pp. 291-312.