

INFORMAL PROOF, FORMAL PROOF, FORMALISM

ALAN WEIR

Philosophy, University of Glasgow

Abstract. Increases in the use of automated theorem-provers have renewed focus on the relationship between the informal proofs normally found in mathematical research and fully formalised derivations. Whereas some claim that any correct proof will be underwritten by a fully formal proof, sceptics demur. In this paper I look at the relevance of these issues for formalism, construed as an anti-platonistic metaphysical doctrine. I argue that there are strong reasons to doubt that all proofs are fully formalisable, if formal proofs are required to be finitary, but that, on a proper view of the way in which formal proofs idealise actual practice, this restriction is unjustified and formalism is not threatened.

§1. Introduction. Recent developments in automated theorem-proving have led some to conjecture that in future all proofs submitted to professional mathematics journals will have to be ‘verified’; for example will have to be in a form which can be machine-translated into a formal proof which can be validated by an automated theorem checker. Thus Freek Wiedijk thinks that improvements in automation will lead to

a quantum leap, and suddenly all mathematicians will start using formalization for their proofs. When the part of refereeing a mathematical article that consists of checking its correctness takes more time than formalizing the contents of the paper would take, referees will insist on getting a formalized version before they want to look at a paper. (Wiedijk, 2008, p. 1414)

whilst Fields medallist Tim Gowers writes:

My own view ... is that ... over the next hundred years or so, computers will gradually be able to do more and more of what mathematicians do ... and eventually supplanting us entirely. (Gowers, 2002, p. 134)

Such claims look reasonable if what has been called the ‘standard view of proof’ is correct:

every mathematical proof should be convertible into a formal derivation in a suitable formal system. (Marfori, 2010, p. 261)¹

However, on the other side of the fence, Yehuda Rav, in an important and influential paper ‘Why Do We Prove Theorems’ Rav (1999), casts doubt on the standard view.

We have what seem like polar opposite views then. One pictures informal proofs as akin to promissory notes, valid only to the extent that they are, or at least can be, backed up by ‘fully formal’ proofs. At the other extreme it is held that informal proofs cannot, in general, be converted or cashed out in formal terms. As well as its relevance to the issues

Received: July 31, 2014.

¹ See also Celluci (2008, p. 1), Goethe & Friend (2010, p. 273) and MacIntyre (2005, p. 2420).

arising from the emergence of automated theorem-proving, this discussion has, of course, especial salience for formalists in the philosophy of mathematics, who seem committed to the ‘standard view’.

In fact there are (at least) two different debates going on here, one over epistemological issues, the other metaphysical. The epistemological debate is a descendant of debates which arose out of the nineteenth century movement for increased rigour in mathematics. This led some, following Frege (but not necessarily with his logicist goals in mind), to set as an ideal in mathematical practice that proofs consist in chains of reasoning with no links missing, no gaps to be filled in by intuitive leaps. What the emergence of formal systems quickly showed, however, was that the huge increase in the length of proofs once they are formalised tends to increase, not decrease, the probability of error, especially when proofs are written out and checked by humans.² But the emergence of automated proof systems raises the possibility that this problem can be overcome. There is now, therefore, a live epistemological standpoint, which I will call ‘enformalism’, and which holds that the goal of having all proofs produced by professional research mathematicians encoded in formal proofs and validated by an automated theorem-checker is both realistic and desirable.

On the other hand, many proponents of doctrines called ‘formalism’ in the philosophy of mathematics are primarily motivated by metaphysical considerations, specifically a desire to avoid commitment to a platonistic ontology and believe that this can be effected either by replacing the notion of truth, as applied to mathematical sentences, by that of proof, or by closely identifying truth with formal proof. I will use ‘formalism’ to characterise metaphysical doctrines of this type. It is formalism the metaphysical doctrine with which I am concerned. The two positions are logically independent. Even if formalism is false, it could be that every mathematical conjecture we ever are likely to consider will in fact be formally provable or refutable in ‘real time’ so that automated verification is feasible. And likewise formalism could be true but enformalism false. We cannot conclude from the fact that a sentence has a formal proof that mathematicians, even using the most powerful computers which are technologically possible, will be able to find and check the proof. Nonetheless there are some looser connections. In particular, the arguments which sceptics about enformalism employ do seem to cast doubt on formalism. My aim in this paper is to defend (metaphysical) formalism by arguing for an account of the relation between informal and formal proof which leaves formalism intact even if scepticism about enformalism is justified.

The structure of the paper is as follows. In the next two sections, I will try to clarify further the informal versus formal proof distinction and tie down a little more carefully which of the many philosophies of mathematics sailing under the flag ‘formalism’ are in the line of fire, if the sceptical line adopted by Rav is correct. In Section 4 I look in more detail at one of Rav’s main targets, what he calls ‘Hilbert’s Thesis’ and in Section 5 give what I hope is a strong version of the argument against it. I then consider a counter-argument, one which proceeds indirectly: Section 6 raises an apparently even more devastating problem for formalistic approaches namely the fact that formal proof far outruns concretely realisable proof. Strict finitism, I argue, is not a coherent response to this problem. Instead I set out, in Section 7, a conception of idealisation which I claim an

² Moreover the process of formalising an informal proof is not, at least at present, a mechanical one and the possibility of an incorrect formalisation leading to a false validation of a supposed theorem cannot be ruled out (cf. Rav, 1999, p. 35).

anti-platonist formalist can legitimately appeal to and which resolves the problems posed by the paucity of concrete proofs. The apparent tangent into strict finitism is necessary for the argument because, I argue in Section 8, the notion of idealisation which liberates the anti-platonist from strict finitism also liberates us from the prejudice that formal proofs must be finitary. I conclude that Hilbert's Thesis, in its finitistic version, is false, and with it the standard view of proof, so that some doubt is cast on enformalist optimism. However by amending the standard view to allow infinitary derivations we arrive at a plausible view of the relation between informal and formal proof. The position, I argue, is eminently compatible with an uncompromisingly anti-platonist formalism.

§2. *Hilbert's Thesis* is Rav's term for the proposition that:

every conceptual proof can be converted into a formal derivation in a suitable formal system (Rav, 1999, p. 11)

By 'formal system' Rav means a purely syntactic set of rules distinguishing those strings or trees of formulae which are 'derivations' from those which are not. It is clear that he also intends that the usual constraints be placed on syntactic systems: wffs and proofs are to be finite, theoremhood recursively enumerable. A 'conceptual proof' by contrast has 'irreducible semantic content' (*ibid*).

Rav is sceptical about the thesis (p. 12 fn. 3, p. 31) declaring it an 'article of faith' (p. 35) rather than something for which plausible grounds have been given. Defenders of the faith still exist, it would seem— 'the still persistent belief in Hilbert's Thesis', (p.28)— but, despite the persistence, 'the number of believers in it is constantly dwindling' (p. 35).

Perhaps so but as we have seen, many take Hilbert's Thesis to encapsulate 'the standard view of proof' and take recent developments in automated theorem checking and proving to provide support for the standard view. If Hilbert's thesis is false then something is amiss with the standard view. Moreover the falsity of Hilbert's thesis would seem to entail the falsity of the philosophical doctrine of formalism, a doctrine which still has its adherents today.

Not that formalism is itself a sharply defined thesis: in philosophy as elsewhere, 'ism' terms tend to have a variety of meanings. In order to sharpen focus, I will exclude certain readings such as the view that mathematical utterances are meaningless and also the doctrine that mathematical propositions lack truth values; though 'formalism' has been used to encompass these extreme doctrines, they will not be considered here. More positively, historical 'formalists' have generally tied mathematical correctness (or truth, if they admitted the notion) to formal provability. Equating, with Rav, 'formal provability' with 'derivability' let us stipulate that a philosophy of mathematics only counts as formalist for present purposes if it rules out the possibility of underivable mathematical truths (and takes mathematical utterances as, in general, having truth values). Finally, in order not to take on too much, let us concentrate on what one might call 'uniform' theories of mathematical language, theories which give the same metaphysical treatment to all parts of mathematics, thus leaving aside Hilbert's own, rather instrumentalistic, formalism, with its completely different treatment of finitary versus infinitary mathematics.

Thus sharpened, formalism seems to entail Hilbert's Thesis (though not conversely). Supposing our system of 'conceptual proofs' is sound, it will generate only mathematical truths; but if formalism is right each such truth is formally derivable. So if Rav's scepticism is about the thesis justified, formalism, as well as enformalism, is in trouble.

It might be said this is of little interest: is formalism in any recognisable sense not philosophically dead? No, there have been signs of revival. Rav himself in a later article

Rav (2007) criticises the views of Jody Azzouni who describes himself as holding to a ‘version of formalism’ (Azzouni, 2004, p. 105) in which ordinary mathematical proofs ‘indicate’ formal derivations. To be sure it is not entirely clear that Azzouni is a formalist in the sense under discussion. The indication relation is left rather open: Azzouni does not claim that the indicated derivations all belong to a single formal system; rather ordinary proofs can indicate derivations from a ‘family’ of formal derivations. Clearly a given body of ordinary proofs cannot indicate both proofs in a formal system with A_1, \dots, A_n as theorems and also proofs in a system with $\neg A_1, \dots, \neg A_n$ as theorems (cf. Rav, 2007, p. 303, fn. 18). So we need to know more about the indication relation and what makes one system, but not a different, inconsistent one, the one which is indicated by a given informal practice. Furthermore Azzouni holds that the indicated derivations need not exist; certainly concrete tokens of them need not exist in the same period as the informal proofs which indicate them. Proofs in ancient Greek geometry indicate 21st century or later derivations. In fact, as he acknowledges, they need not ever exist, they may be too long ever to be written down (Azzouni, 2006, p. 154). And yet it is these non-existent proofs which are supposed to explain the consensus among mathematicians about which informal proofs are correct! Azzouni himself seems later to have doubts about the plausibility of all this:

I kept falling (against my will) into a view that mathematicians had to be engaged in *something* like sophisticated syntactic pattern-recognition while perusing informal mathematical proofs, so that they would be sensitive (without realizing it) to a background of *nonexistent* formal derivations. (Azzouni, 2009, p. 25)

moving to an ‘inference package’ view of mathematical reasoning. Thus it is no longer clear that he remains a formalist.

Even if Azzouni is no longer a formalist, and perhaps never was, there are other contenders for contemporary defenders of formalist positions, for example among the fictionalists. Of course the fictionalist, at least of the most straightforward sort, does not hold that mathematical theorems express *truths*, *a fortiori* not truths linked in some tight sense with provability. For according to such a fictionalist all existential mathematical theorems are false, the only mathematical truths are trivial truths, for example universal generalisations of the form $\forall x(\phi x \rightarrow \psi x)$, ϕx expressing a mathematical property hence, according to the fictionalist, true of nothing.

However this can mislead. The fictionalist does have a distinction between correct and incorrect mathematical sentences, utterances and assertions. ‘There are infinitely many primes’ is a correct mathematical claim, ‘there are infinitely many complex solutions to each quadratic equation’ an incorrect claim. Fictionalists tempted by deflationary views of truth will presumably be tempted to hold that it is true that there are infinitely many primes, even though they must reject the idea that what makes it true is the existence of an infinitude of abstract objects. What does make it true? That it is part of the ‘story of mathematics’ as ‘Oliver Twist was born in London’ is, in some sense, part of Dickens’s story (Field, 1989, pp. 203)? If the implication is that to be true is to be proclaimed, with or without proof, a theorem in the mathematical literature and never have any challenge to that claim accepted by the mathematical community then we should reject this version of fictionalism. There will be no end of mathematical claims, some true, some false, which never make it into the mathematical literature at all, never being considered by actual mathematicians. And given the number of “theorems” whose purported proofs were later discovered to be faulty, we can be fairly sure some falsehoods will remain accepted for all time as “theorems”.

If, however, the fictionalist declares that what makes a mathematical sentence correct/true is the existence (or possible existence) of a proof, then this counts as a version of formalism.³ The fictionalist, however, now incurs the burden of proof, as it were, borne by all formalists: to tell us what proofs are, to tell us whether they are concrete tokens or abstract objects, to tell us what it is about the meaning of mathematical statements that enables non-synonymous statements about the existence of proofs and disproofs to provide their truth-conditions and to respond to the many objections raised against formalism.

Further contemporary formalists perhaps include Gabbay (2010), though Gabbay writes ‘Contrary to traditional (game) formalism, my proposal shall not involve an attempt to provide formal derivations of each and every arithmetic truth’ (p. 221). Weir (2010) is explicitly formalist in identifying what makes a mathematical truth true with the existence (perhaps unknown) of a proof with, that is, its provability.

So formalism as a philosophy of mathematics is not dead. Moreover, turning from the metaphysical aspect to the epistemological, if Marfori and the others cited in footnote 1 are right then the standard view of informal proof links such proofs to formal derivations as required by Hilbert’s thesis and thus renders enformalism plausible. There is evidence for their claim in the increasing numbers taking seriously the computerised formalisation effected by programs such as *Coq*, *Isabelle* and *Mizar*, in the trend among some category theorists, such as those attracted to the univalent foundations program, towards a practice of informal proof which can be readily computerized (Awodey, 2010, p. 3) and in explicit statements such as the remarks by Wiedijk and Gowers cited at the beginning of this paper.⁴

In sum, Hilbert’s thesis is of interest in that a compelling argument against it seems to yield in turn a compelling argument against a live metaphysical position in philosophy of mathematics and also a serious challenge to a view of the relation between informal proof and formal derivation which has significant support from mathematicians. Has Rav, then, presented a compelling critique of Hilbert’s Thesis?

§3. Firstly, let us look at Rav’s distinction between conceptual proof and formal derivation, between

a conceptual proof of customary mathematical discourse, having an irreducible semantic content,

which is to be distinguished from a

derivation, which is a syntactic object of some formal system. (Rav, 1999, p. 11.)

Derivations he defines fairly standardly:

a (linear) derivation in a formalised theory T is a finite sequence of formulas in the language of T, each member of which is either a logical axiom, or an axiom of T, or is the result of applying one of the finitely many explicitly stated rules of inference to previous formulas in the sequence. With some minor modification one similarly defines a tree derivation. *ibid* (see also Rav, 2007, p. 301.)

³ An alternative, suggested to me by Mary Leng, is to take correctness to be determined not by formal provability from axioms but by ‘logically following’ from axioms, where the notion of logical consequence is defined in terms of a primitive notion of modality. This retains some distance between formalism and fictionalism.

⁴ See <http://www.mizar.org/project> for *Mizar*.

But Rav goes further than merely distinguishing semantics from syntax, saying:

Once we have crossed the *Hilbert Bridge* into the land of meaningless symbols, we find ourselves on the shuffleboard of symbol manipulations, ... these symbols do not encode meanings ... it is the very purpose of formalisation to squeeze out the sap of meanings in order not to blur focusing only on the logico-structural properties of proofs. Meanings are now shifted to the metalanguage, as is well known. (*op. cit.* p. 12.)

As the axioms in a formalised theory are strings of symbols, syntactic objects, the issue of their truth is vacuous (*op. cit.* pp. 18–19.)

Rav here seems to be subscribing to the view that because symbols *are not* meanings therefore *they cannot have* meanings, or cannot be assigned them. But this is wrong-headed, as Quine says:

There is a short-sighted but stubborn notion that a mere string of marks on paper cannot be true, false, doubted or believed. Of course it can, because of conventions linking it to speech habits and because of neural mechanisms linking speech habits causally to mental activity. (Quine, 1995, p. 94.)

Bearing that caveat in mind, I will stick to Rav's informal proof versus formal proof/derivation distinction generalising the latter notion slightly to:

A derivation in a system S is a member of the intersection of every set containing a base set of one-step proofs—axioms— and closed under the formal rules for generating larger proofs out of smaller which characterise S.

This is the usual idea of defining formal systems by induction. What is crucial is how we understand 'formal rule'. The usual gloss is along the lines of: 'a rule which appeals to nothing outside the system'. This is vague but greater precision is unlikely to be had in respect of a concept as fundamental as this (the concept of 'formal rule' not itself being part of formalised mathematics). We can help clarify by looking to applications of the idea. Thus in grammar we assume we start from atomic symbols whose internal structure, if any, is utterly irrelevant and characterise those sequences which are complex well-formed expressions without appeal to relations the sequences may have with any entities outside the strings of symbols. And similarly, taking certain formulae as our base elements and dubbing these 'axioms', complex proofs are specified as structures, strings, trees or whatever of wffs, without bringing in relations to anything else. In diagrammatic derivations we must similarly have formal rules making no appeal to anything outside the diagrams which determines how to build more complex diagrammatic derivations out of simpler.

In the simple case where derivations are linear strings of wffs, each later string follows precisely from the earlier ones by one of the rules, with a similar determination of the output of proof-constructing rules in more complex architectures; a derivation is a *gapless* proof, in other words. An important point to note is that there can be derivations in natural languages such as English, even more obviously when they are augmented with a little mathematical notation. Note that I did not require that the formal rules of a derivation be what we would recognise as *logical* rules. Quite obviously, when learning the school rules for addition and multiplication, we are not given axioms from which we can derive, using the machinery of *Principia Mathematica* perhaps, $7+5=12$. But we are given formal rules

in the above sense, just as we are when taught one or other of the various algorithms for long addition or multiplication (so a long multiplication set out in canonical form constitutes a derivation). Similarly, I would guess that the overwhelming majority of students being taught calculus are not exposed to axioms but rather taught rules, e.g. basic rules for differentiation, power rules, product rules, quotient rules, the chain rule and so on. These are completely formal rules.⁵

These sort of examples show that *derivations abound*. There are concrete tokens of mathematical derivations all over the place, in textbooks, exam scripts, worked examples in engineers' rough notes, in the hardware of computers computing values for various functions and so on. The derivations in textbooks and notebooks often take the form of a linear string of formulae each derived from earlier ones by one of the rules of the system or by applications of identity laws such as Leibniz' law. There are also partial derivations, derivations with gaps; these are derivation sketches in a definite and concrete sense. They could, if we could be bothered, be expanded to full derivations. By this we do not mean that if we kept going without let up or error for a few billion years or more a full derivation would result. Rather they could be completed in a very ordinary every day sense of *could*: with a few hours more of patient writing out, we could produce a gapless proof.⁶ Indeed Rav's 'fully rigorous' proof (2007, p. 305) that there is a unique left and right identity in a group is easily expanded to a derivation of that form without having to dress it up in a Hilbert-style axiomatisation.

Rav, to be sure, might quibble about whether derivations, as construed above, are formal proofs, for he seems to have in mind a model of proof in which all the non-logical action, as it were, is loaded into the axioms. But, at least when our background logic has the requisite structural rules, the choice of whether to use rules, or axiomatic versions of rules, is essentially a matter of convenience or aesthetics. Thus to take some simple cases, any system S with an axiom scheme ϕ is equivalent to one without the axiom scheme but with the schematic rule (supposing a sequent architecture) $T \vdash \phi$, T a truth constant. Conversely a relatively simple sequent rule of the schematic form,⁷

$$\begin{array}{l} \text{from } P_1, \dots, P_n \vdash Q \text{ conclude} \\ R_1, \dots, R_m \vdash S \end{array}$$

can be dropped in favour of the axiom scheme

$$[(P_1 \rightarrow (P_2 \rightarrow \dots (P_n \rightarrow Q) \dots))] \rightarrow [(R_1 \rightarrow (R_2 \rightarrow \dots (R_m \rightarrow S) \dots)]$$

given \rightarrow I (conditional proof), \rightarrow E (modus ponens) and sufficient structural rules.

But whether the mathematical reasoning we take to be non-logical is to be formalised axiomatically or in rule form is not the key point for sceptics about formalisation. Rather it is whether it can be formalised at all, whether axiomatically or in rule form.

⁵ As a referee points out, one might well call these sorts of examples *calculations* rather than *derivations*. Even if there is this intuitive distinction between those categories, the important point is they are both sub-species of a more general category, that of a formal system, as characterised above. For convenience I will continue to use 'derivation' for the more general category.

⁶ Or, more and more commonly nowadays as noted at the outset, we might write the sketch up in a form which we can input to an automated theorem-prover which might then give us fairly quickly, and in some cases in readily human-readable form, the full derivation.

⁷ omitting the constants which will occur in the schematic letters in any non-trivial rule;

§4. For none of the previous points about ‘really existing’ derivations directly touch on Rav’s main case, namely that the proofs which professional mathematicians are interested in, the sort of things which appear in mathematics journals (at least in advance of the promised era of machine-generated formalisations), are not derivations, nor partial derivations; indeed they are not, in general, remotely like derivations (or ‘calculations’, if you prefer) such as a simple chain of equations in calculus. How then does Hilbert’s thesis link the informal proofs of the professional mathematician with derivations of the humdrum type discussed at the end of the last section? I suggest the following expanded version of the thesis as a way of making sense of the possibility of such links:

Hilbert’s Thesis II: In any cogent mathematical practice there is a systematic process of transformation (not necessarily known to the practitioners) which turns any correct proof into a (suitably related) finite derivation in a formal system *S*. The system *S* in question is determined by the informal practice and its transformation process; in particular, the formal rules of *S* are rules which are implicit in the mathematical practice.

Note that it is entirely open what the concrete tokens of formal systems are, whether sounds, ink marks, electronic states of computer systems, biological compounds or whatever. The notion of formal system also applies, as the observations of the previous section made clear, both to natural and formal languages. Not only that, there is no restriction solely to linguistic systems— diagrammatic reasoning can be formal as well. This is clear in systems such as Peirce’s diagrammatic proof-architecture, where there are precise rules for what counts as a legitimate proof-step. The status of other diagrammatic modes, ancient Euclidean demonstrations or modern category-theoretic diagrams for example, is more problematic. They perhaps stand to Peirce-style pictorial derivations as informal proof does to linguistic derivations.

But though there is this openness to a wide range of formal systems, for Hilbert’s Thesis II (HTII) to have definite content we must still require that it be a determinate matter what formal system underlies a given area of mathematics in the sense that provability⁸ in that system is at least a necessary condition for the correctness of the suitably transformed informal proof. Does that mean that any formal system can validate the correctness of a given body of informal practice? No, we will at least require that the system be non-trivial. Are there any further constraints? For a radical pluralist about logic, perhaps not; however those who believe there are some semantic constraints on soundness of logical rules will require that correct informal proofs transform into derivations in which the rules for the logical operators are sound. At the other extreme from the radical pluralist, a logicist will hold that informal proofs will be transformed into derivations in which all axioms and inferences are logical, governing formal correlates of operators deemed to be logical.

Clearly HTII still leaves a lot up in the air about the relationship between informal proof and derivations. We need to fill in more detail about what it is for the formal system to be ‘suitably related to’ and ‘determined’ by informal mathematical practice, work which would require more extensive investigation than can be given here. But as an example of one way one might proceed, take inferentialist accounts of the meaning of logical and mathematical operators. If these accounts are right, then there will be syntactic rules which capture the actual reasoning practice of competent mathematicians. A formalisation

⁸ which I will equate, throughout, with ‘existence of a proof, known or unknown’;

of these rules, then, will count as a suitably related system determined by the informal practice. However, even if inferentialism is true, enformalism will not directly follow. For our inferential practice, like our linguistic practices in general, may not be completely transparent to us. If it is sufficiently opaque then the enformalist goal of having all informal proofs encoded in formal, machine-validated proofs may not be practically feasible. Nonetheless the inferentialist picture in tandem with HTII makes enformalism at least plausible; the combination is also music to the ears of the metaphysical formalist, though of course the problem of the ontological status of the formal derivations remains to be resolved.

As it happens, I do not myself believe inferentialism is correct. The link between meaning and linguistic practice, including inferential practice, is more complex than inferentialism allows. But the stronger the case for holding that there is some determinate (at least for the most part) connection between a subject's grasp of logical and mathematical operators and her tacitly following formal rules involving those operators, the stronger the case for enformalism and formalism. Of course even if this is so, the epistemology of the relationship between informal and formal proof will remain complex. To illustrate, even if inferentialism is correct and I have a full grasp of logic and mathematics, and even if there is a formal proof of C from P_1 to P_n using rules I implicitly accept, it does not in the least follow that any old informal chain of reasoning I produce which takes us from the premisses P_i to C is a good proof. Similarly, even if there is a proof of Goldbach's Conjecture from the axioms of PA, simply affirming the axioms and then uttering 'hence every even number is the sum of two primes' does not constitute a good proof, indeed any sort of proof.

What then is the connection between informal proof and formal proofs, if HTII is correct? Can we set this out in a way which explains why good proofs increase our mathematical knowledge and understanding? These are very difficult questions in the epistemology of mathematics. Although we can rule out trivial cases— one-step proofs consisting of no more than appending an unobvious consequence to axioms and so forth— providing a more substantive answer is a major challenge. It seems too restrictive, for example, to require that a correct informal proof have, at some interesting level of abstraction from detail, the same overall structure as a formal derivation which, on this picture, validates it as correct. We should, though, contra Azzouni, insist that the moves used in the formal derivation should be moves featuring in the practice of the informal provers, or at least moves they would recognise and acknowledge.

However these tricky and important questions are not the topic of this paper, bearing, as they do, primarily on the epistemology, not the metaphysics, of proof and truth in mathematics. But whatever view one takes on what makes an informal proof of a truth P of a given mathematical discipline a good proof, if HTII is correct then the formalist can give an answer to the question what makes P a truth, namely the existence of a formal derivation of the proposition expressed by P using axioms and rules implicit in the informal practice of exponents of that discipline.

§5. But is HTII correct? One class of objections Rav raises to his original formulation of Hilbert's thesis concerns the absence of formal axioms for large chunks of mathematics (and, one could just as well add, the absence of formal, non-logical, inference rules too).

No axioms have ever been proposed for even a fragment of matrix theory. It is a typical unaxiomatised theory. What is then the logical status of its extensive results? (Rav, 1999, p. 16)

Rav also mentions graph theory, measure theory and number theory (as opposed to axiomatised Peano arithmetic) as unaxiomatised theories. This may pose a problem for enformalism.⁹ But the point is not fatal to formalistic approaches so long as it is plausible that there is, implicit in mathematical practice, rules (and perhaps axioms never implicitly brought to the surface) which, when formalised, generate derivations of every truth of matrix theory, measure theory etc. Even if there is some fuzziness as to which formal system is implicit, even if it is not completely determinate whether it is system S , or S' or S'' , this can still be accommodated if pretty much everything we hold to be a determinate truth of the theory is provable in each of the reasonable 'precisifications' of informal practice with no determinate falsehoods provable in all.

A much more interesting argument of Rav's, with respect to formalism as a metaphysical picture, goes as follows. Suppose we are trying to understand an informal proof of B from A and we break it down into a linear sequence of claims:

$$A \rightarrow A_1, A_1 \rightarrow A_2, \dots, A_n \rightarrow B$$

(the arrow representing merely an informal path from antecedent to consequent). Baffled initially by the move from A to A_1 we eventually see how the author must take it to go and explain it to a student by interpolating A' yielding $A \rightarrow A' \rightarrow A_1$.

But the process of interpolations for a given claim has no theoretical upper bound. In other words, how far has one to analyse a claim of the form 'from property A , B follows' before assenting to it *depends on the agent*. There is no theoretical reason to warrant the belief that one ought to arrive at an atomic claim $C \rightarrow D$ which does not allow or necessitate any further justifying steps between C and D . This is one of the reasons for considering proofs as *infinitary objects*. (Rav, 1999, pp. 14–15, italics in the original).

For present purposes, Rav's suggestion that proofs are *infinitary* objects is of most interest. It should be noted, though, that his motivation is not in the metaphysical issues which the formalist addresses but much more in the epistemology of proof. Putting it crudely and oversimplifying, he argues that proofs matter more in mathematics, even with regard to generating knowledge and understanding, than the theorems proved. Informal proofs yield new concepts, methods and techniques.¹⁰ Moreover comprehension of proofs, he emphasises, is agent-relative, dependent on background knowledge.

Why, though, should that lead to the conclusion that proofs are infinitary objects? If, for example, HTII was true, and if, contrary to the sceptical suggestion above, the formal inferential rules implicit in our practice are not opaque to us then we would have a theoretical reason to suppose that by transforming an informal proof sufficiently far we would arrive at 'atomic' moves which every competent agent, regardless of background, would see as warranted outright and in need of no further justification. But Rav's argument suggests

⁹ Though it is not hard to imagine formalising matrix theory by taking a formalisation of complex analysis and adding some axioms and rules to govern manipulation of matrices conceived of as ordered sets of complex numbers.

¹⁰ Methods for doing what, one might ask? Generating new proofs and so further concepts, methods and techniques without end? It is hard not to feel that at least one intrinsic goal of mathematical methods and techniques is settling the truth of interesting mathematical theses by establishing them as theorems.

a metaphysically interesting reason why one might contrapose here, and conclude to the falsity of HTII.

For Rav throughout assumes that derivations must be finite; in fact he takes for granted the usual view on which the theorems form a recursively enumerable set. Call such systems, ‘Gödelian systems’. Granted this, we can put the objection to HTII in the following way. We know from Gödel’s limitative results (augmented by Rosser) that many important mathematical theories, when formalised as Gödelian systems, are negation-incomplete: any consistent extension of the arithmetic theory Q for example. But we also know (it is generally held) that any Gödel sentence $G: \forall x \neg (\text{Bew}_T(x, \ulcorner G \urcorner))$ is true, where ‘ Bew_T ’ represents provability from a true (in the standard model) extension T of Q , T being the formalisation determined by the transformation process of HTII as the one underlying, let us suppose, our arithmetical practice at a given time. ‘Provability’ here is a recursive notion; that is, x proves y is a recursive relation over codes of sentences and derivations, thus $\exists x(x \text{ proves } y)$ is a recursively enumerable relation. But then given that T is true (hence consistent), no informal proof of T ’s Gödel sentence can be transformed into a formal derivation, subject to the Gödelian constraints, since there is no such derivation (in a formalisation of theory T). And since we have insisted that there has to be a definite formal system fixed by every mathematical practice (at least where we assume each sentence has a determinate truth value), the fact that the Gödel sentence for the theory T which formalises current practice is formally derivable in an extended system $T \cup \{G\}$ is irrelevant since G is not an axiom of T or any reasonable idealisation of our arithmetical practice. Moreover given the tight tie in formalism (most simply outright identification) of truth with provability in an underlying formal system S , the truth of G seems to refute that philosophy of mathematics.

But informal proof, the argument here seems to go, is different. Rav seems to be thinking of appeal to consequence relations, such as full second-order consequence, which go beyond finitary, first-order r.e. provability:

the implicit underlying logic of mainstream group theory is second-order logic. (op. cit. p. 17)

The argument against HTII can then be presented in the following way. If we interpret ‘derivation’ to mean derivation in a system in which the theorems are recursively enumerable, then in many interesting mathematical theories there will be statements¹¹ such that neither they nor their negations are derivable, in the appropriate formalisation. Yet in many cases, we will firmly believe we know that one or other of the statement or its negation is true, not through some mystical intuition but because we feel we have an (informal) proof of it. Hence informal proof \neq formal derivation.

§6. We have here, then, a sharp version of the dilemma for the standard view of informal proof and of course for formalists, a dilemma generated by Gödelian results. If truth = formal provability then undecidable sentences must lack a truth value. It follows that many Π_1 sentences of arithmetic lack truth value, including some we think we know are true on the basis of informal proofs of them. Even if most ‘ordinary’ mathematical theorems escape this fate, even if *Mizar*, *Coq* and co. can validate these with a formal derivation, this is still a most implausible result and leaves a potential infinity of intuitively

¹¹ including ‘natural’ statements of interest to mathematicians in general, such as have been investigated by H. Friedman, J. Paris, L. Harrington, L. Kirby and others.

determinate mathematical theses, some perhaps informally provable, lacking in truth values and formal derivations or refutations.

Nonetheless though I think this argument does spell deep trouble for HTII, it does not dispose of formalism (though it may dispose of the ‘standard view of proof’). In order to explain why, I need to look at an even more basic problem for formalism, at least for those formalists (surely the majority) who are motivated by suspicion of abstract objects.

For a standard objection to formalism is that the analysis of the truth of mathematical statements as consisting in, or somehow grounded in, the provability of strings does nothing to support a metaphysical view of a world devoid of abstract objects because the strings which are proven, and the proofs which prove them, are themselves abstract objects. The only way to get round this problem, the argument continues, is to identify syntactic objects with concreta, with mereological fusions of concrete tokens or some such. But since there are surely only finitely many of those, (at least if they are to be graspable in an unproblematic sense by us) we are plunged into a uncompromising strict finitism, and for most¹² this constitutes a *reductio ad absurdum* of formalism.

What really causes the difficulty here is the fact that the shortest derivations of mathematical theorems are often orders of magnitude larger than the theorems themselves. If that were not so, a really determined formalist/strict finitist could coherently hold that there are only finitely many mathematical sentences, all shorter than some upper bound in size, but each provable or refutable, challenging the non-finitist to come up with an instance of a mathematical truth outside the bounds. If the bounds are set high enough, the non-finitist will not be able to provide a concrete refutation of the strict finitist thesis.

But we know from Gödelian speed-up considerations that for many a provable sentence, the shortest derivations of it or its negation are vastly longer than the sentence itself. For example, for any polynomial f , given a system S^{n+1} of $n+1$ th-order arithmetic, there will be theorems of the system which are expressible in the language of n^{th} order arithmetic and whose shortest proofs (including the theorem itself, as last line) in S^{n+1} are l long but whose shortest proofs in n th order arithmetic S^n are $f(l)$ long. If S^k is the highest order system we ever use, there will be S^k truths, expressed in short, graspable sentences, whose proofs contain more steps than the estimated number of neutrinos in the universe, truths for which there can be no concrete proof in other words. Quite simple examples of this sort of thing can be given, for example Neil Tennant’s primality claim:

$$(2^{2^{2^{2^{2^2}}}} + 1) \text{ is prime}$$

(Tennant, 1997, p. 152) which may well be such that no concrete proof or refutation exists, even though the problem is a simple one in Δ_0 arithmetic.

It is clear that the formalist has to appeal to idealisation of our concrete practice. One approach to idealisation is via what one might call ‘supernaturalised epistemology’, that is appeal to ideal beings who can do what is only, in the obscure and problematic phrase, “in principle possible” for us. I have argued that appeal to such beings is inadmissible for a naturalist and that this route is a dead end (Weir, 2010, pp. 181–191). A better strategy emerges when we reflect on one key reason why idealisation is essential even for those sympathetic to the strict finitists’ concretist predilections, namely the all-pervasiveness of abbreviation. Although concrete tokens of proofs and derivations are not only finite but

¹² Though not all: the mathematician Doron Zeilberger defends a strong finitist position he calls ‘ultra-finitism’; Edward Nelson also expressed strong finitist leanings.

quite small (at least if they are to be digestible by humans), this does not mean there is a small upper bound on the size (as abstract strings or trees) of wffs or proofs for which a concrete token can exist. If I am writing out derivations for a fragment of arithmetic in a language L_T which has only successor and zero signs ‘S’ and ‘0’ I might tire of writing out numerals SSSSSSSS0 etc. in full and introduce metatheoretic abbreviations such as $S^{10^{1000}}0$ to stand for the numeral consisting of ‘0’ prefaced by 10^{1000} occurrences of ‘S’. There is no upper limit to what might be ‘reached’ by abbreviation here, given the open-endedness of notational innovation.

Let F' be an *abbreviational system* for F if a) F' conservatively extends F by the addition of rules which permit uniform substitution of occurrences of P' for string P (of the same grammatical category) occurring in formulae in proofs in F and b) P' is shorter (as a string) than the corresponding P .¹³ Of course here we speak with the platonist. The formalist is therefore under an obligation to show how such assertions in proof theory can be true without ontological commitment to abstract objects. If this can be done, however, she is entitled to point out that many theorems for which there are no concrete derivations in unabbreviated systems have concrete derivations in extended abbreviatory systems.

Note that an ‘abbreviatory’ concrete token such as:

$$S^{10^{1000}}$$

is as much a token of an abstract string as this one: SSSS0. We can think of it as a stretching of the equiformity relation among tokens. Letters in different fonts can look rather different; and this written token– ‘ten’– is a very different physical entity from the vocal utterance of that word. Similarly ‘SSSS0’ and S^40 look a bit different but we can treat them as tokens of the same type. What abbreviation enables us to do is produce tokens which are equiform with no concrete unabbreviated token, none we can grasp anyway.

Abbreviation applies not only to expressions and wffs but also to derivations. This is a token:

\emptyset (1)	$\forall x \forall y (Sx = Sy \rightarrow x = y)$	Axiom
1 (2)	$S^{10^{1000}}0 = S^{(10^{1000})-1}0$	Hyp
\emptyset (3)	$S^{10^{1000}}0 = S^{(10^{1000})-1}0 \rightarrow$ $S^{(10^{1000})-1}0 = S^{(10^{1000})-2}0$	1, $\forall E \times 2$
1 (4)	$S^{(10^{1000})-1}0 = S^{(10^{1000})-2}0$	2,3 $\rightarrow E$
\emptyset (5)	$S^{(10^{1000})-1}0 = S^{(10^{1000})-2}0 \rightarrow$ $S^{(10^{1000})-2}0 = S^{(10^{1000})-3}0$	1 $\forall E \times 2$
1 (6)	$S^{(10^{1000})-2}0 = S^{(10^{1000})-3}0$	4,5 $\rightarrow E$
...	As [5-6] for $2 \times (10^{1000}-3)$
1 (2×10^{1000})	$S0 = 0$	
1 ($[2 \times 10^{1000}]_+1$)	$\exists x (Sx = 0)$	$(2 \times 10^{1000}), \exists I$
\emptyset ($[2 \times 10^{1000}]_+2$)	$\neg \exists x (Sx = 0)$	Axiom
1 ($[2 \times 10^{1000}]_+3$)	\perp	$([2 \times 10^{1000}]_+1), \dots +2, \neg E$
\emptyset ($[2 \times 10^{1000}]_+4$)	$S^{10^{1000}}0 \neq S^{(10^{1000})-1}0$	$([2 \times 10^{1000}]_+3), \neg I$

of a $(2 \times 10^{1000})_+4$ long derivation. At any rate, if we add tokens of an explanation of ‘As [5-6] for $2 \times (10^{1000}-3)$ ’ steps to the token proof we have complete instructions which enable us to figure out (working in metamathematical proof theory) exactly what

¹³ See, for example, the extended Frege systems of Cook & Reckhow (1979, sec. 4).

the abstract derivation is. This is not a derivation sketch. It is not to be compared to a proof by cases in which one writes out, say, two of the eight cases and leaves the other six as exercises. Here we have an explicit recipe for fixing on the right abstract object, just as much as in a token proof of absurdity from say $SSS0 = SS0$ in which we write out every step without abbreviation.

True, one could not grasp the above token proof without grasp of decimal notation and exponentiation yet the underlying formal system itself is just arithmetic in a language L_T with 0, S as constants. We can imagine someone being trained to understand L_T and carry out unabbreviated derivations in that language yet not grasping decimal notation nor having exponentiation defined for them. So surely the above token derivation is not a token of an L_T -proof?

Certainly there is a difference between abbreviated and unabbreviated derivations and someone may grasp an unabbreviated derivation but not the abbreviated one. The same can be said of informal proofs vis à vis formal. This does not alter the fact that, for someone who *does* grasp the abbreviated derivation, the token derivation may be as compelling as an unabbreviated one. In fact, one utility of abbreviated derivations lies in the fact they will often be much more compelling than unabbreviated derivations, where, that is, the latter actually exist; the other key advantage, then, is that they enable some theorems to be concretely derived which could not otherwise be. Nor does the above observation, concerning what we might call the *conceptual surplus* of some types of abbreviation, alter the fact that the token in the extension of T (provided it is conservative) demonstrates (though not, of course, with absolute, Cartesian-sceptic proof certainty) a fact about *provability in L_T* , not provability in some other system, e.g. of decimal arithmetic.

We have here an insuperable problem for strict finitism. Abbreviation means that in the corpus of our mathematical utterances, derivations and proofs, many wff and proof tokens are ‘stranded’ high up in a hierarchy of complexity. It is not possible— not feasible for the real flesh and blood humans who engage in the practice and understand its results— to construct and grasp tokens of all the wffs and derivations of lower complexity. We can see this with $S^{10^{1000}}$ 0. No matter what notational innovations we come up with, we will never at any time have a system in which tokens, abbreviated or unabbreviated, of all the numerals for numbers $< 10^{1000}$ are instantiated. This means that lots of concretely realised expressions will be such that some (in many cases most) of their constituents will not be concretely realised. Similarly lots of concrete derivations, such as the one above, will be “missing” many or most of their sub-proofs, namely those which are not concretely realised.

This does not prevent the strict finitist from giving a physicalistic specification of concrete surrogates for wff and proof (cf. Goodman & Quine, 1947; Weir, 2010, chap. 6). But it means that the specification cannot be given in the usual inductive fashion, as the intersection of all inductive sets containing the base set and closed under the complexity-forming operations. And this in turn means we cannot prove even very simple facts about wffs and proofs— about unique readability, expressive adequacy and so on; nor define, by recursion for example, even fairly simple concepts such as the degree of complexity of a formula. Inductive proof and recursive definition are simply unavailable to the strict finitist; the chain from simplest atoms to formulae and derivations of arbitrary complexity is broken. But a philosophy of mathematics which requires us to give up inductive proof and recursive definition is absurd.

§7. Thus the realities of actual mathematical practice ‘do in’ strict finitism. But far from posing an insuperable problem for formalism, the above considerations show how to

defuse the argument against formalism from HTII. With regard to the latter, remember that the formalism I am considering here does not maintain that mathematical practice is ‘mere symbol-shuffling’; it holds rather that mathematical utterances (including metamathematical utterances) generally express truth-valued assertions. The formalism need not come in the specification of the sense or literal content of mathematical sentences but in explaining what makes them true or false. To help make things clear I will utilise the SMALL CAPS convention (Horgan, 1994, p. 316), ‘THERE ARE MOUNTAINS’ is used to assert the mind-independent existence of mountains, ‘There are numbers’ to make an existential claim whose truth is determined not by the world being the way it is represented to be, if we read the literal content in the same way as in the mountain claim. Rather, it is determined by the existence of a derivation of (a translation) of ‘there are numbers’. Of course a good deal of the hard work required to defend this variety of formalism lies in making sense of this distinction between literal content, which in general has nothing to do with proofs or derivations, and the derivations which make-true or make-false, and I am not going to attempt that large task here.¹⁴ But if we grant for now the possibility that such a position can be made out, can it escape the fatal defects of strict finitism?

Yes: just as the formalist believes that THERE ARE NO NUMBERS, but also believes there are infinitely many prime numbers, so too the formalist believes that THERE ARE NO ABSTRACT SYNTACTIC TYPES, but also that there are infinitely many wff types and proof types, inductively generated from base sets in the usual way. Granted that formalism can make sense of mathematical utterances in general as truth-valued assertions, the formalist can define meta-mathematical concepts applicable to formal systems in the usual way, for example the formal concepts of provability and of truth (for a narrower object language). Alternatively the concepts may not be explicitly defined but introduced via a formal theory, for example an axiomatic theory of truth which can be used to explain how the truth and satisfaction conditions of complex expressions depend on those of simpler ones.

This means that, by contrast with the strict finitist, the formalist can, without bad faith, actually state the problem raised in the previous section, namely the existence of a whole class of sentences of decidable fragments of mathematical languages which are true but whose formal proofs are too long to be concretely realised. But the tools which the formalist possesses and which enable her to state the problem enable her to solve it: the formalist claim is that for every mathematical truth there is a derivation of that truth (not that THERE IS A CONCRETE PROOF); for every mathematical falsehood there is a refutation. And in mathematical logic textbooks THERE ARE CONCRETE DERIVATIONS, or at least proof sketches, of those metamathematical Π_2 “lower-case” facts (for every truth, there is a proof ...) linking formal truth and formal derivation for such sub-languages as Δ_0 arithmetic.

Is this too good to be true, too easy a victory for the formalist? Well, what would it be for the formalist position to be false? Formalists, of course, are no more immune from *mathematical* error than anyone else. If the formalist affirms a sentence S (of formal system F) but in fact THERE IS A CONCRETE DERIVATION (IN F) OF $\neg S$, then the formalist has made a mathematical error.¹⁵ But the considerations of the previous section, the argument against strict finitism in particular, show us that affirming S though THERE IS NO CONCRETE STRUCTURE WHICH, SUITABLY INTERPRETED, CONSTITUTES A DERIVATION (IN F)

¹⁴ This is attempted in Weir (2010, pp. 25–38)

¹⁵ If F is inconsistent there is simply no truth or falsity to be had in the system, at least if provability distributes over the connectives of the formal system just as truth does.

OF S, does not mean that the formalist has spoken falsely, even by formalist lights. For there may BE¹⁶ A CONCRETE FORMAL PROOF that there is an abstract derivation of S. Even if one does not wish to call this a derivation-in-F it plays a similar epistemic role as a concrete derivation in F. The case is similar to, though a little more complex than, an abbreviated derivation of S in a conservative extension of F. We might call the F*-derivation of \ulcorner there is a derivation-in-F of S \urcorner an indirect derivation-in-F. An indirect derivation-in-F of S counts as making S true (as a sentence in system F, or one whose truth is determined by system F) just as much as a derivation in F would. So the concrete truth-maker, if one wishes to use this language, of a sentence S of system F can take, on this view, one of two forms: a derivation-in-F of S, or an F*-derivation of the proof-theoretic claim that there is an F-derivation (for suitable F*).

If, however our formalist affirms S (hence \ulcorner S \urcorner is true but also affirms that THERE IS NO CONCRETE STRUCTURE WHICH CAN PLAY THE ROLE OF A DERIVATION IN F* THAT THERE IS AN ABSTRACT F-PROOF OF S, for any suitable F*¹⁷ then, indeed, the formalist would be guilty of inconsistency. But formalism is not an inconsistent position, for the formalist need not, should not, make any such affirmation.

But what about the concretely undecidable claims which the formalist takes to be determinate, for example the primality claim P above? Neither the appeal to indirect derivation nor the ubiquitousness of abbreviation resolves the problem of concrete indeterminables. No matter what abbreviatory conventions we introduce and however much we appeal to indirect derivations, there will always be concrete utterances which we can neither prove— whether directly or in abbreviated form— nor disprove. (The point of the appeal to abbreviation is to rule out strict finitism as an anti-platonist response to the problems formalism faces not to attempt to provide concrete derivations, direct or indirect, for every mathematical truth.)

Once we see, however, that the formalist is as entitled to affirm the truth of metamathematical results as results in arithmetic or set theory, we also see that she can consistently and legitimately affirm $\text{True}(\ulcorner P \urcorner) \vee \text{True}(\ulcorner \neg P \urcorner)$, where P is our concrete indeterminable. Given transparency of truth, which is unproblematic in these non-paradoxical contexts, the formalist must then as a consequence affirm $P \vee \neg P$ and so will be convicted of error if THERE IS NO CONCRETE PROOF OF $P \vee \neg P$ or at least no proof that there is an abstract proof of this instance of LEM. But even if the background logic is non-classical (as I think it should be) simply including all instances of LEM as axioms of the formalised arithmetic theory will ensure that there are very short concrete proofs backing up all these assertions of instances of LEM. Hence no inconsistency.

The sceptic will counter-reply that the inconsistency lies in the combination of a plausible account of the meaning of the connectives together with the formalist views on the truth/derivation connection, that is in an incoherent combination of claims about the recursive structure of truth and its inter-relations with the recursive structure of formal proof. If disjunction and negation, at least in mathematical contexts, are truth-functional

¹⁶ As always in mathematics, the copula here is timeless; moreover there is no requirement that the EXISTENCE of the concrete structure ever be known. But it has to be a structure which we flesh and blood humans could, had we known of it, (in a perfectly ordinary sense of ‘could’) interpret (by suitably interpreting its elements and the structural inter-relations) as a derivation in a formal system we use. By using only a relatively non-theoretical, quotidian notion of ‘could’ there is no naturalistically suspect importation of modality here of the sort to be found, some claim, in modal structuralism.

¹⁷ ‘Suitable F*’ raises some issues I will look at in the next section.

operators then since $P \vee \neg P$ is true either P is or its negation is (P our concrete undecidable as above). Since “truth is determined by concrete provability”, the formalist is guilty of incoherence in affirming, as she must, that THERE IS NO CONCRETE PROOF of P (nor a concrete proof there is an abstract derivation of P) nor a concrete proof of its negation (nor that there is an abstract refutation).

The mistake in this argument lies in the scare-quoted “truth is determined by concrete provability”. As we have seen, there are no interesting systematic claims about the syntax or semantics of the concrete corpus of our mathematical utterances. Generalisations about the syntactic and semantic structure of the formal idealisation of our concrete output can only be made from within a similarly formal metatheory. The formalist, in such a theory, makes the claim that the *formal idealisation of truth* coincides with the *formal idealisation of provability*. But this claim can be backed up by a *concrete proof* in the metatheory of arithmetic that Δ_0 is decidable and that true = provable = the existence (not: ‘EXISTENCE’) of a proof.¹⁸ At no point need the formalist affirm something which she has reason to believe will lack a concrete proof. The objection to formalism fails.

The upshot, then, is that formal idealisation of informal proof is treated as similar in some important respects to idealisation in empirical science. For example, just as one can view a continuous distribution of real values representing mass density in a fluid as the image of an injection of physical properties and relations into a structure which may be richer— perhaps fluids are large but finite assemblages of elementary particles each of uniform mass density— so too the injection of the finite body of our actual mathematical utterances into formal systems idealises by injecting a finite structure into a smoother, more manageable, infinite one. The statement that this injection exists is made true, as with all other mathematical statements, pure and applied, not, even in part, by the EXISTENCE OF ABSTRACT OBJECTS but by its provability. (Of course large questions are raised for the formalist here, both about the purposes of idealisation in mathematics— how similar is it to idealisation in empirical science?— and also about whether the formalist is entitled to the view that there can be truths of applied mathematics, pace Frege, 1903. See Weir, 2010, chap. 5.)

However there is a further problem which the formalist faces which an absolutist platonist does not. The formalist criterion for asserting or assenting to a mathematical sentence S which is being uttered as part of a practice whose optimal formalisation is F should be, if the considerations above are correct:

THERE IS A CONCRETE STRUCTURE which can play the role of a derivation in F of S or a derivation in F^* that there is an abstract F -proof of S , for any suitable F^*

where F^* need not be F .¹⁹ But if there are no constraints on F^* , if it can be a trivially inconsistent theory for example, then everything is correctly assertible. We can prove ‘The string $\ulcorner 0=1 \urcorner$ is provable from the axioms of PA’ from a trivial F^* . We thus need to impose the condition that any such F^* be consistent. But that is not sufficient. If the literal content of sentences provable from F^* is to tell us something about abstract proofs in F , we must be

¹⁸ Or at least backed up by a concrete sketch that there exists such a formal proof in an extension of the language of arithmetic in which notions of truth and proof are introduced.

¹⁹ And where, when S is a disjunction, there is no implication that THERE IS A CONCRETE DERIVATION, in F or F^* , of either disjunct, even where *formal provability* distributes over the connectives, even where the theory is prime in other words.

able at the least to *define* in F^* , formal-proof-in- F (or at any rate axiomatically characterise it) in a way which captures our intuitions about proof-in- F . If F^* cannot do this, or if it proves too many counterintuitive results about F -provability then we will view the literal content of strings in the F^* -system as telling us not about *formal proof* in F , but some other, if perhaps related, notion— *schproof* in F or whatever. (Likewise, as noted, a string ‘ $0 = 3$ ’ in a formal system of arithmetic modulo 3^{20} does not express the same proposition as that string in standard arithmetic, it expresses a truth of modular arithmetic not arithmetic.) The strong formalist, then, will not be committed to the assertibility of palpable absurdities in arithmetic, analysis or whatever. Will all indeterminacy be eliminated though? Can we rule out the possibility of two different systems F^* and F^{**} which are optimal candidates for formal theories of proof-in- F but which disagree on the provability of some non-obvious conjecture P ? The answer is no, as we will see when I utilise the response to strict finitism given above to address the worries from Rav we started from, the concern in particular that formalising proof is a process of interpolation which ‘has no theoretical bound’ because informal proofs are ‘infinitary objects’ (Rav, 1999, pp. 14–15).

§8. The considerations adduced thus far in favour of idealising concrete proof give the strong formalist entitlement to assert that truth distributes over the connectives just as proof does, and hence justify affirming LEM even in the case of concrete indeterminables— but only for negation-complete fragments of language such as Δ_0 arithmetic. What of the rest? What of Gödelian incompleteness? Is our informal proof of the truth of each Gödel sentence with respect to *every* true extension of Q not a case of an *infinitary object* as Rav says, one which does not admit of reduction to a formal proof in a single system in which theoremhood is recursively enumerable?

Here I agree with Rav in giving an affirmative answer, though not, perhaps, in a manner which he would accept. For Rav explicitly assumes that formal proofs are finite. Why think that? Why think that formal proof— derivability— must be recursively enumerable? What reason is there to think that because the structure which justifies a theorem is infinitary the proof cannot be formalised?

I hope it will be clear by now that the knee-jerk reaction which most philosophers will give— proofs are epistemic constructions which have to convince finite creatures like ourselves— is hopeless. To be sure, the concrete proofs which convince us are finite; not just finite but actually quite small, much, much, smaller, in their number of symbols say, than the estimated number of electrons in the observable universe. But we do not think this means that our idealised notion of formal derivation should not permit strings longer than, say 10^{1000} in length. Idealisations must preserve the key feature, ‘recursiveness’, of our informal practice, the feature which enables us to grasp new unencountered sentences and proofs. But this is ‘recursiveness’ solely in the sense of grasp of iterable rules whose output can in turn be fed in as input to the self-same rules. Nothing in that notion of recursiveness rules out iteration into the transfinite.

Just as we have this small graspable token

1	(1) $\neg\{2^{2001}\}: a=b$	Hyp.
1	(2) $\neg\{2^{2001} - 2\}: a=b$	1 DNE
	2^{1000} -2 steps of DNE
1	$(2^{1000} + 1) a=b$	2^{1000} DNE.

²⁰ One in which we choose to use ‘=’ rather than ‘ \equiv ’ for congruence modulo n .

of an unfeasibly long proof, (where $\ulcorner \neg\{n\}P \urcorner$ abbreviates P prefaced by n negation signs) so too this

\emptyset	(0) $\neg\text{Bew}(0, \ulcorner G \urcorner)$	Given.
	$\dots \dots \dots$	
\emptyset	(n) $\neg\text{Bew}(n, \ulcorner G \urcorner)$	Given.
	$\dots \dots \dots$	
\emptyset	(ω) $\forall x(\neg\text{Bew}(x, \ulcorner G \urcorner))$	$n \in \omega$; ω rule.

is a small, graspable token of an $\omega+1$ long proof, using the ω rule, of a sentence G which is a Gödel sentence with respect to a finitary partial idealisation of informal arithmetic, such as Q plus standard first-order logic. Note that these tokens are not concrete realisations of schematic proofs. Provided one accepts the infinitary mathematics involved in setting up the notation, there is no more ambiguity in specifying which formal object the token picks out than there is in knowing which term ‘ $S^{10^{1000}}$ ’, picks out in a formalisation of the language L_T (given that one grasps exponentiation).

Nor does the above requirement restrict us to countable infinitary logics, to $L_{\omega_1\omega_1}$. To take a simple example, if we have a language with \beth_{\aleph_0} singular terms, a language closed under conjunctions and disjunctions less than the first inaccessible $> \beth_{\aleph_0}$ in size, then

$$\bigwedge (t_\alpha = t_\alpha) \ \alpha < \beth_{\aleph_0}$$

is a small token, abbreviating a rather long conjunction, a token which is perfectly comprehensible to anyone who understands enough transfinite arithmetic and the idea of indexing singular terms by ordinals, and which again unambiguously specifies a formal correlate.²¹

Just as axiom sets and inference rules stated within a particular proof-architecture can be ordered in terms of proof-theoretic strength, so too the formal frameworks themselves can be ordered in terms of power. Polynomially-bounded logic is a weaker framework than decidable logic (if $P \neq NP$), which is weaker than first-order predicate logic, and so on through countable infinitary logics, accessible logics (which debar wffs and proofs from being of inaccessible cardinal size) and beyond. A policy of accepting only minimal idealisations of concrete practice, idealisations which allow for iteration of wff-forming and proof-forming operators but are otherwise as close as possible to actual human capacities, would surely drive us down to polynomially-bounded logic, if not below. This is a price too high to pay, and once we see that there is no non-arbitrary cap to place on formal frameworks in terms of cardinal size, we should let \aleph_α flowers bloom, for arbitrarily high α .

To be sure, moving beyond the weaker notions of proof opens up the possibility, noted in the previous section, that two different formal systems F^* and F^{**} could equally well idealise the notion of proof in F, yet come to different views as to whether $\vdash_F S$. The formalist response to this should be the same as the response to the differences between ordinary and modular arithmetic. The notion of formal proof or derivation differs between the two systems so that $\vdash_F S$ can be true* in the F^* system and false** in F^{**} : ‘ $\vdash_F S$ ’ is ambiguous just as ‘ $3 = 0$ ’ is. This may lead, then, to some ‘hyper-indeterminacies’ in that there is no unequivocal sense in which certain types of sentence– the Continuum Hypothesis, the Axiom of Choice perhaps– are provable/true even as relativised to a particular infinitary formal system. But for an anti-platonist, this should not be a surprising or disconcerting development.

²¹ For the formalist, of course, to say that there is a formal correlate for concrete token c is not to say that THERE IS A FORMAL CORRELATE.

Of course allowing idealisations of our concrete proof-theoretic practice to encompass infinitary proof, as well as arbitrarily long finite proofs, will not in and of itself ensure negation-completeness. Although the weak first-order theory Q is indeed negation-complete, in the context of first-order logic augmented by the ω -rule, there are infinitary systems of set theory in which e.g. the Continuum Hypothesis is undecided, as well as some in which it is provable, others in which it is refutable. The formalist should say (speaking from within a given meta-framework) that the standard formalisations of the hypothesis mean somewhat different things in each formal framework (since it is the frameworks of rules and axioms of proof which give meaning to the symbols of the system) and that the different propositions expressed are respectively, i) not determinately true nor false in the undecided case, ii) determinately true in the system in which it is provable and iii) determinately false in the other. This is not counter-intuitive in the way that denying truth value to sentences of arithmetic (and thus to sentences of the most basic sorts of proof theory) is; on the contrary, it is perfectly consonant with an anti-platonist perspective.

§9. Conclusion. I have addressed the question whether Rav is right to challenge Hilbert's Thesis that every 'conceptual proof' can be converted into a formal derivation and if so what the implications are for formalism. My conclusion is that if one assumes formal proofs are finitary objects, Rav's scepticism about Hilbert Thesis I (and II) is justified. That assumption, however, is itself unjustified. Any coherent philosophy of metamathematics must allow a role for idealisation and a plausible conception of that role leads to a liberalised notion of formal derivation. If we alter the 'standard view of proof', if such it is, by dropping the finiteness requirement then the amended Hilbert Theses lose their plausibility.²² Moreover on this conception of proof and the relationship between informal and formal proof, the tight connection between truth and proof envisaged by the formalist is not subject to the usual criticisms based on limitative theorems such as Gödel's incompleteness theorems.

§10. Acknowledgments. The author is grateful to members of the audiences at presentations on this topic at seminars and conferences at the Universities of Aberdeen, Bristol, Cambridge, Edinburgh, Glasgow (special thanks to Gareth Young here), Manchester, Oxford, Queen's University Belfast, St. Andrews, Stirling, Stockholm, Trinity College Dublin and at a joint APA/ASL session on formalism in San Diego. My thanks also to Catarina Dutilh Novaes and an anonymous referee for very helpful comments.

BIBLIOGRAPHY

- Awodey, S. (2010). Type theory and homotopy. Archived at <http://arxiv.org/abs/1010.1810>.
 Azzouni, J. (2004). The derivation-indicator view of mathematical practice. *Philosophia Mathematica*, **12**, 81–105.
 Azzouni, J. (2006). *Tracking Reason: Proof, Consequence and Truth*. Oxford: Oxford University Press.
 Azzouni, J. (2009). Why do informal proofs conform to formal norms. *Foundations of Science*, **14**, 9–26.

²² Nor is infinitary proof a no-go area for automated proof search and checking, using codings of course, as the study of recursive ω rules shows, cf. Baker *et al.* (1992).

- Baker, S., Ireland, A., & Smaill, A. (1992). On the use of the constructive omega rule within automated deduction. In Voronkov, A., editor. *International Conference on Logic Programming and Automated Reasoning*. Berlin: Springer, pp. 214–225.
- Celluci, C. (2008). Why proof? what is proof? In Corsi, G. and Lupacchini, R., editors. *Deduction, Computation, Experiment. Exploring the Effectiveness of Proof*. Milan: Springer, pp. 1–27.
- Cook, S., & Reckhow, R. (1979). The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, **44**, 36–50.
- Field, H. (1989). *Realism, Mathematics and Modality*. Oxford: Blackwell.
- Frege, G. (1903). *Grundgesetze der Arithmetik II*. Jena: Pohle. Reprinted in 1962, Hildesheim: George Olms. Translated in 2013 by Ebert, P. A., & Rossberg, M. as *Basic Laws of Arithmetic*. Oxford: Oxford University Press.
- Gabbay, M. (2010). A formalist philosophy of mathematics part I: Arithmetic. *Studia Logica*, **96**, 219–238.
- Goethe, N., & Friend, M. (2010). Confronting ideals of proof with the ways of proving of the research mathematician. *Studia Logica*, **96**, 273–288.
- Goodman, N., & Quine, W. V. O. (1947). Steps towards a constructive nominalism. *Journal of Symbolic Logic*, **12**, 97–122.
- Gowers, T. (2002). *Mathematics: A Very Short Introduction*. Oxford: Oxford University Press.
- Horgan, T. (1994). Naturalism and intentionality. *Philosophical Studies*, **76**, 301–26.
- MacIntyre, A. (2005). The mathematical significance of proof theory. *Philosophical Transactions of the Royal Society*, **363**, 2419–2435.
- Marfori, M. A. (2010). Informal proofs and mathematical rigour. *Studia Logica*, **96**, 261–272.
- Quine, W. V. O. (1995). *From Stimulus to Science*. Cambridge, Massachusetts: Harvard University Press.
- Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, **7**, 5–41.
- Rav, Y. (2007). A critique of a formalist-mechanist version of the justification of arguments in mathematicians' proof practices. *Philosophia Mathematica*, **15**, 291–320.
- Tennant, N. (1997). *The Taming of the True*. Oxford: Clarendon Press.
- Weir, A. (2010). *Truth Through Proof – A Formalist Foundation for Mathematics*. Oxford: Oxford University Press.
- Wiedijk, F. (2008). Formal proof – getting started. *Notices of the American Mathematical Society*, **55**, 1408–1414.

PHILOSOPHY, SGOIL NAN DAONNACHDAN
 OILTHIGH GHGLASCHU\UNIVERSITY OF GLASGOW
 GLASGOW, G12 8QQ, SCOTLAND
 E-mail: Alan.Weir@glasgow.ac.uk