Correctness, Artificial Intelligence, and the Epistemic Value of Mathematical Proof

James Owen Weatherall

Department of Logic and Philosophy of Science
University of California, Irvine

Jesse Wolfson

Department of Mathematics
University of California, Irvine

Abstract

We argue that it is neither necessary nor sufficient for a mathematical proof to have epistemic value that it be "correct", in the sense of formalizable in a formal proof system. We then present a view on the relationship between mathematics and logic that clarifies the role of formal correctness in mathematics. Finally, we discuss the significance of these arguments for recent discussions about automated theorem provers and applications of AI to mathematics.

1. Introduction

That a mathematical proof be correct is neither sufficient nor necessary for it to have epistemic value. By "correct", here, for the sake of the argument, we mean that the proof can be formalized, i.e., translated into a formal proof system. (Below, we will call this sense of correctness "formal correctness" and distinguish it from "mathematical correctness".) The context for this sense of "correct" is the recent literature on the Standard View of mathematical proof (Azzouni, 2004; Avigad, 2006, 2008, 2021, 2022; Weir, 2016; Hamami, 2018, 2022; Tatton-Brown, 2023), which holds that ordinary "informal" proofs as generated by working mathematicians and published in high quality mathematics journals, should be seen as describing, or encoding, or abbreviating, or "sketching" (in the words of Saunders

Mac Lane (1986, p. 377)) a formal proof that could, at least in principle, be generated from the informal proof (though whether this process is "routine" (Mac Lane, 1986) or an additional "creative act" (Avigad, 2021; Larvor, 2022) varies among proponents of the view). By having "epistemic value", we mean being well-suited to play the various epistemic roles that proof does, and should, play in mathematical practice. In other words, epistemic value is what makes proofs "work", as proofs, for the purposes that mathematicians seek to attain. We will say more below about what these roles are.

Our thesis may seem striking or even outrageous. Of course mathematical proofs should be correct! And if a proof cannot be formalized, surely that is a signal that something odd or surprising is going on. We grant this—as, we think, would most contributors to the literature on the Standard View. But even so, there is a long history of debate over the status of the Standard View, arguably originating in debates between figures such as Hilbert, Brouwer, and Poincaré in the early 20th century, with recent entries in the philosophical literature from Rav (1999), Detlefsen (2008), Cellucci (2008), Larvor (2012), and Tanswell (2015) criticizing aspects of the Standard View.[1] (We will discuss some of these other arguments in more detail below.) These arguments tend to have one of two forms: either they emphasize that formal proofs are purely syntactic, whereas informal ones seem to have some further content related to knowledge of the subject matter, so that they are not *merely* abbreviations of formal proofs; or else they emphasize that there are significant mismatches between what appears to be required for a successful formal proof and what is needed for an informal one.

The present paper should be seen as part of the same tradition. But we wish to put our emphasis in a different place—and to offer what we believe is a deeper critique of the Standard View, grounded in a form of anti-logicism that we think is prevalent in many areas

---

[1] A rich and parallel discourse in the mathematics community runs through Polya (1954), Lakatos (1976), Thurston (1994), Hersh (1997b), Manin (1998), Arnold (2006) and Harris (2019, 2022).

of mathematics, but which we have not seen clearly expressed or defended in the philosophical literature. On our view, formal correctness is secondary to the goals and epistemic value of proof. More, the formalizability of mathematical proof flows from a different sort of activity than the ones that mathematicians are engaged in when they develop and present informal proofs. Proofs are formally correct, when they are, not because correctness is the normative ideal governing mathematicians' activities, nor because correctness underwrites the epistemic value of those activities. Instead, informal proofs are formally correct because logicians have developed a rich and successful mathematical theory of mathematical practice intended to capture various aspects of mathematical argument. Formal correctness tracks mathematical correctness because formal correctness was designed to do precisely that.

In other words, the formal systems in which formal correctness is established have been invented and tuned with the goal of making existing informal proof methods formalizable. Formalizability speaks to the success of this activity in mathematical logic—but it is irrelevant to the question of what makes the informal proofs epistemically valuable. Insofar as failures of correctness signal problems with proofs, it is not because formal correctness is the goal of the proof; rather, it is that failures of formal correctness indicate salient differences in the reasoning between the (formally) incorrect proof and other examples of successful mathematical reasoning.

Although our motivations for criticizing the Standard View are similar to those of previous authors, we feel the issue has become increasingly urgent. In recent years, we have seen rapid advances in the areas of computer-assisted theorem proving and proof verification. Those working in these areas often invoke the ideology of the Standard View to explain and justify their goals: if it is correctness that is the ultimate goal, then automated systems that can ensure correctness are the ideal proof generators. This sort of reasoning has led to some striking soothsaying: in the future, one should expect AI mathematicians to outpace human mathematicians at proving theorems, in much the same way that AI game players

have outpaced their human competitors in Go, chess, Jeopardy, and other games (Bory, 2019). Many groups have already devoted substantial resources to realizing this vision of future mathematics; and some very prominent mathematicians have become boosters for AI-driven mathematics (e.g. Buzzard (2022); Gowers (2022); Tao (2024)).

We do not want to gainsay the future value of AI for mathematics. We, too, are impressed by the rate of progress in recent years, and we suspect that computer assistance will change many aspects of mathematical life—just as Mathematica and Matlab changed the lives of undergraduate calculus students and many applied mathematicians a generation ago. But we maintain that if one accepts our principal theses, then the mere generation of correct proofs is not, in itself, a contribution to mathematics, at least not in the absence of several other necessary conditions. The broader mathematical community needs to be clear about what these sorts of activities can and cannot contribute, and what risks are posed to more traditional mathematical activities by reorienting towards machine-driven methods. While we will not attempt to give an exhaustive list of what additional criteria are jointly sufficient for proofs to have epistemic value, we will point below to several conditions that are needed.

The remainder of the paper will proceed as follows. We will begin by reviewing the Standard View and some previous criticisms of it to better situate what we are trying to do here. In the following section, we will elaborate on what we mean by the "epistemic value" of a proof, and use that machinery to argue that correctness cannot be sufficient for a proof to have epistemic value. We will then turn to the other half of our claim, and argue that correctness is not even necessary, strictly speaking. Here we will offer several examples of cases in which incorrect proofs (or attempted proofs) turned out to be of exceptional epistemic value. Then, in section 6 we will return to the view sketched above of how logic should be seen as related to mathematics, and we will offer an account of why one might have thought formal correctness was necessary. We will conclude by reflecting on how our arguments bear on the question that motivated us, concerning what sort of contributions

AI systems can make to the mathematical enterprise.

2. The Standard View: For and Against

The Standard View of mathematical proof is a claim about the relationship between mathematical proof as encountered in the wilds of mathematical practice and formal proofs.[2] One of the Standard View's most prominent defenders describes it as follows:

> When someone in the mathematical community makes a mathematical claim, it is generally possible to express that claim formally, in the sense that logically adept and sufficiently motivated mathematicians can come to agreement that the formal claim expresses the relevant theorem. One justifies an informal claim by proving it, and if the proof is correct, with enough work it can be turned into a formal derivation. Conversely, a formal derivation suffices to justify the informal claim. So an informal mathematical statement is a theorem if and only if its formal counterpart has a formal derivation. Whether or not a mathematician reading a proof would characterize the state of affairs in these terms, a judgement as to correctness is tantamount to a judgment as to the existence of a formal derivation, and whatever psychological processes the mathematician brings to bear, they are reliable insofar as they track the correspondence. (Avigad, 2021, p. 7379)

In other words, a proof plays a justificatory role in establishing a theory only if the proof can be translated into a formal derivation; and insofar as a proof is "correct", it *can* be so translated.

The Standard View comes in several flavors. Azzouni (2004), for instance, defends what he calls the "derivation-indicator view of ordinary mathematical proof" (p. 85), which is the view that the role of proof in practice is to convince other mathematicians of the existence of a certain kind of "derivation", or formal proof—and not, necessarily, to directly encode that proof. This is a version of the standard view that elaborates the reasons – pragmatics, efficiency, expedience – that mathematicians give informal proofs, even though

---

[2]The present review is intended only to give a sense of the literature and motivate the arguments that follow. For an extended introduction and exhaustive literature review, see Tanswell (2024).

it is ultimately formal proofs that matter. For Azzouni, the Standard View has important explanatory virtues: in particular, it explains why there is intersubjective agreement among mathematicians about which proofs are correct. It is because these proofs are the ones that successfully indicate the existence of a formal proof.

The Standard View of proof is closely connected with another view, also sometimes called the Standard View, concerning *rigour*. As Hamami (2022, p. 409) puts it, "Mathematical proof is the primary form of justification of mathematical knowledge. But in order to count as a *proper* mathematical proof, and thereby to function *properly* as a justification for a piece of mathematical knowledge, a mathematical proof must be *rigorous*." He goes on to describe (and defend) what he calls the Standard View: "According to this view, a mathematical proof P is *rigorous* if and only if P can be *routinely translated* into a formal proof (p. 410)." For our purposes, the differences between these versions of the Standard View will not matter. We see "(formally) correct", "rigorous", "proper", and other similar terms as success terms indicating that a proof has achieved its key epistemic aims, and therefore has epistemic value in the sense described above, i.e., that it is able to play the epistemic role in mathematical practice that proof is expected to play. And so we take the Standard View to be the claim that the epistemic value of an "informal" proof is underwritten by the in-principle existence of a correct formal proof that bears some relationship ("formalization") to it. Our arguments are intended to respond to this broad family of views.

The Standard View thus described is hardly new. Some authors identify its source in Frege, though it is not clear if he truly held a version of the Standard View because he maintained every inference has some "non-formal" component. But there are good reasons to think that both Hilbert and members of his school subscribed to the Standard View. Cellucci (2008), for instance, calls the view the Hilbert-Gentzen Thesis (c.f. Hilbert, 1928; Gentzen, 1964), and states it as follows: "Every real proof can be represented by a formal proof." Hamami (2022) highlights articulations of the view by Mac Lane (1986) and Bour-

baki (1970), though he also argues that it is broadly descriptively accurate of 20th century mathematics, post Hilbert. It is also endorsed by some contemporary mathematicians. Very recently, versions of the Standard View appear to be behind some mathematicians' calls for a significant investment in automated proof checking (e.g. Hales (2008); Grayson (2017); Scholze (2020); Buzzard (2024)).

Criticisms of the Standard View can also be found throughout the mathematical literature, at least implicitly. These remarks most often appear in works where mathematicians reflect on their own methods and discipline, such as the now-classic remarks from Fields medalist William Thurston, best known for his work on low dimensional topology and geometry, on "Proof and Progress in Mathematics" (Thurston, 1994), and work by Reuben Hersh (1993, 1997a,b). But the philosophical literature opposing the standard view is most often traced back to Rav (1999).[3] Rav presents a thought experiment wherein we entertain a mathematical oracle that could rule on the truth or falsity of any mathematical proposition. He suggests this would realize a Hilbertian formalist's vision of mathematical utopia—and that it would simultaneously spell doom for mathematics.

He goes on to argue that, "the essence of mathematics resides in inventing methods, tools, strategies and concepts for solving problems which happen to be on the current internal research agenda or suggested by some external application. But conceptual and methodological innovations are inextricably bound to the search for and the discovery of proofs, thereby establishing links between theories, systematising knowledge, and spurring further developments" (6). In other words, proof is an essential part of mathematics because the activity of trying to produce, refine, and improve proofs is often the context for mathematical innovation and deeper understanding. But the formal correctness of proof is only incidental to this role in mathematical practice, for two reasons. First, it is very

---

[3]Rav (1999) itself sits downstream from a much longer running discourse which includes Thurston (1994); Hersh (1997b); Lakatos (1976); Polya (1954); Poincaré (1905–1906), as does the present article.

rarely within a fully formal setting that mathematicians develop significant new ideas—and of course, automated theorem provers are not the kinds of things that can have new ideas at all. Secondly, he argues, formal proofs are some distance removed from proofs in the sense that Rav thinks proofs fruitfully contribute to mathematics, because formalizing a proof involves breaking the link to the intended semantics of the informal mathematical arguments. We are no longer talking about groups, or manifolds, or locales; instead, we are manipulating symbols according to certain rules. For these reasons, the Standard View misidentifies both why proofs are important, and what features of proofs most contribute to that importance.

In the 25 years since Rav's essay appeared, others have also mounted attacks on the Standard View. Detlefsen (2008), for instance, emphasizes that "Mathematical proofs are not commonly formalized, either at the time they're presented or afterwards. Neither are they generally presented in a way that makes their formalizations either apparent or routine" (p. 17), suggesting that formalization is not a central concern for the mathematicians generating proofs, and that it is not essential to, or even especially closely related to, establishing or exhibiting rigor. Easwaran (2009) extends this by proposing "transferability" as a key criterion, i.e. that the "the point of mathematical proof is to give the reader a means to come to their own autonomous knowledge without relying on the testimony of the author, rather than any particular connection to anything formal."[4] Tanswell (2015) presents a different argument, which is that in general an informal proof can be formalized in many different ways, and often in different systems, suggesting that it is a mistake to hold – as some, such as Azzouni (2004) and perhaps Avigad (2021), apparently do – that an informal proof is just a compressed way of signaling that a (particular) formal proof exists. Instead, the many-to-one relationship of informal proofs to formalizations suggest that formal proofs represent different ways of elaborating or commenting on the structure of an informal proof, and not

---

[4]Personal correspondence with authors.

vice versa. Or consider Larvor (2022), who, responding specifically to Avigad (2021), argues that formalists often conflate two different visions of what an informal proof is: either an informal proof is itself a syntactic object (and thus just a poor version of a formal proof), or it has some semantic content. If it is the former, then it looks like the Standard View is trivial; if it is the latter, there is a significant gap between what informal proofs actually accomplish and what a formal proof can hope to recover.

3. Formal Correctness is not Sufficient

We now turn to the principal arguments of this paper. We will offer two arguments that correctness is not sufficient for a mathematical proof to have epistemic value.[5] The first argument concerns additional necessary conditions on the assertions proved; the second concerns additional necessary conditions on the proofs themselves.

We begin with the most obvious: the fact that a proof is correct does nothing to guarantee that the proposition established by the proof is *interesting*. By interesting we mean what working mathematicians mean when they describe a result as interesting, namely, that they learn something from it that is relevant to their goals. We claim that a theorem must be interesting to mathematicians working in some area or other for it to have epistemic value. How could a theorem fail to be interesting? One way is for the assertion to be trivial, in the sense that no one learns anything new from the result. Another way in which a proof could fail to be interesting is that the assertion proved concerns structures or properties that are not well-integrated with the rest of mathematics, so that even true theorems about them do not have any bearing on mathematical practice. The proposition proved is not relevant to mathematicians' goals. And so on.

One might be reminded, here, of the famous proof that there are no uninteresting (real)

---

[5]These arguments are similar to those of Rav (1999) and others, though below we will draw connections to other mathematicians' commentary on how they see their contributions, which we think enrich the arguments.

numbers. For suppose there were an uninteresting number. Take the set of all such numbers. By the Well-Ordering Theorem, there exists a well-ordering on that set and thus a unique least element. But surely *that* number is interesting! One might argue likewise about uninteresting but correct proofs: proofs are themselves rich mathematical objects, and any given proof will have features that distinguish it from others. The core of this response is to maintain that "interest" in our sense is not necessary for epistemic value. Surely epistemic value is in the eye of the beholder, and a proof that may not be of interest to mathematicians trained to work on certain things and beholden to current fashions may nonetheless have some epistemic value in other contexts—perhaps, even, by virtue of the role it plays in computer assisted proof generation.

But this response misses the joke, and the point. Epistemic value, here, concerns the positive epistemic contribution that proofs play in realizing the goals of actual mathematical practice. And whatever one might take the goals of mathematics to be, surely generating sequences of sentences allowed by some proof system or other should not be counted among them. Mathematicians do both more and less than this. The results of rote proof generation will not, in general, contribute to the epistemic activity of mathematicians.

What are mathematicians doing, beyond simply generating correct proofs? A key goal of working mathematicians is to investigate and ultimately understand mathematical structures. Very often, the most important work in this direction has nothing to do with proof, so much as identifying definitions that both clearly and adequately capture mathematicians' ideas, and also lead to fruitful new work elaborating on what follows from those definitions in ways that provide insight into the structures mathematicians aim to describe. One might reasonably worry that we are changing the subject. The standard view is about proofs, whereas we are now discussing definitions, proof techniques, and so on. But this is precisely to the point. When mathematicians speak of "proof", there is an ambiguity between the practice of generating proofs, which involves articulating definitions, making conjectures,

developing techniques, and so on; and there are the artifacts that result at the end. Our claim is that the epistemic value of proof should be located in the rich range of activities that working mathematicians engage in to produce their artifacts, not in the artifacts themselves. Simply generating correct artifacts is not enough.

On aspect of this perspective is nicely captured by the Russian algebraic geometer Yuri Manin, who writes,

> Mathematicians have developed a very precise common language for saying whatever they want to say. This precision is embodied first of all in the definitions of the objects they work with, stated usually in the framework of a more or less axiomatic set (or category) theory, and in the skillful use of metalanguage (which our natural languages provide) to qualify the statements. All the other vehicles of mathematical rigor are secondary, even that of rigorous proof. In fact, barring direct mistakes, the most crucial difficulty with checking a proof lies usually in the insufficiency of definitions (or lack thereof). In plain words, we are more deeply troubled when we wonder what the author wants to say than when we do not quite see whether what he or she is saying is correct. The flaws in the argument in a strictly defined environment are quite detectable. Good mathematics might well be written down at a stage when proofs are incomplete or missing, but informed guesses can already form a fascinating system: outstanding instances are A. Weil's conjectures and Langlands's program, but there are many examples on a lesser scale. (Manin, 1998, p. 166)

Definitions are where mathematicians identify what they find interesting.[6] Without the right definitions, proofs are unlikely to be of epistemic value.

Of course, what follows from good definitions is also important, and that is established by proofs. We do not mean to say that proofs do not matter. But we would argue that what makes proofs epistemically valuable is that they show not only *that* some assertion or other is true, but also *why* it is true, by showing how it follows, using accepted techniques, from prior results that one already understands and accepts. Now, one can argue that some proofs are more explanatory than others, or that some theorems of interest may not admit of proofs that

---

[6]At least some leading early formalizers were also vividly aware of this, e.g. Huntington (1937, pp. 493-494).

are clear and digestible to humans, and yet merely knowing if they are true is of value because of the role they play in proving yet other claims. But none of this contravenes the fact that all else being equal, simple, perspicuous, and comprensible proofs are of special value; and proofs that establish a theorem without being understandable to *any* human mathematician do not contribute to the activity mathematicians are engaged in. A necessary condition for a proof to have epistemic value is that it contributes to mathematicians' understanding of the structures they seek to study—specifically, mathematical structures that are believed to be well-integrated with the rest of mathematics, and ultimately, with the natural and social worlds.

Taken together, these arguments are meant to establish that correctness is not sufficient because correct proofs do not necessarily establish interesting propositions in ways that contribute to mathematicians' understanding of their subject matter. These arguments are not intended to be novel: to the contrary, our goal here is to isolate the core of ideas that are ubiquitous among mathematicians, and frequently expressed by leading figures. Take, for instance, Thurston:

> [W]hat [mathematicians] are doing is finding ways for people to understand and think about mathematics.
>
> The rapid advance of computers has helped dramatize this point, because computers and people are very different. For instance, when Appel and Haken completed a proof of the 4-color map theorem using a massive automatic computation, it evoked much controversy. I interpret the controversy as having little to do with doubt people had as to the veracity of the theorem or the correctness of the proof. Rather, it reflected a continuing desire for human understanding of a proof, in addition to knowledge that the theorem is true.
>
> On a more everyday level, it is common for people first starting to grapple with computers to make large-scale computations of things they might have done on a smaller scale by hand. They might print out a table of the first 10,000 primes, only to find that their printout isn't something they really wanted after all. They discover by this kind of experience that what they really want is usually not some collection of "answers"—what they want is understanding. (Thurston, 1994, p. 162)

Thurston goes on to argue that an "...emphasis on theorem-credits has a negative effect on mathematical progress" (Thurston, 1994, p. 172) because it obscures both the importance of understanding as an output of mathematical practice and the importance of understanding acquired from others in producing proofs in the first place.

As with the quote from Manin, this passage from Thurston is just one well-known example of a prominent mathematician expressing this sort of view. But we need not multiply examples to make the point, since we find that it is widely recognized even among advocates for the Standard View. Take Avigad (2022, §§3-5), for instance, who develops an account of mathematical understanding and mathematical depth, and acknowledges both that mathematicians place great value on understanding and also that formal proofs, or at least, some prominent computer generated ones, apparently do poorly on delivering understanding. (He also acknowledges the importance of good definitions.) But he goes on to argue that perhaps there is a different kind of understanding that at least some computer-assisted proofs can provide, even if it is not quite the kind of understanding mathematicians have typically sought out. We are happy to grant that this can happen. But it is not an argument that formal correctness is sufficient for a proof to have epistemic value. At best, it establishes that there may be multiple ways in which proofs can have epistemic value.

4. Formal Correctness is not Necessary

We now turn to showing that correctness is not necessary for a mathematical proof to have epistemic value. This direction is more subtle, and we find fewer precursors in the literature. Consider that on a common strategy to establish that one property is not necessary for another, one wishes to exhibit, or at least show the existence of, something that instantiates the second property but not the first. In the present case, that would mean presenting a proof that has epistemic value but which is not formalizable. Of course, this is possible. As even the defenders of the Standard View freely admit, real, published mathematical proofs

often contain errors, and as such they are not amenable to formalization. And yet they may nonetheless have considerable epistemic value: the errors may be trivial, so that when corrected, the proof is formalizable; or the proof may present a successful proof strategy even though it does not accomplish its goals as implemented; or the erroneous proof may have a gap, identifying which advances the field in some way; or the proof may introduce new techniques that turn out to be fruitful elsewhere.

One might justly complain about this argument, however. First, proofs with simple errors fail to be formalizable in a trivial way (and, indeed, Avigad (2021) argues that standard mathematical methods are effective in part because they are robust against small errors, so that erroneous proofs are effectively formalizable). Meanwhile, to cite "proofs" that have more serious errors, but which contribute to mathematical knowledge in other ways, is arguably question begging. "Proof", here, should be understood as a success term. An argument with significant errors, ones that are not easily corrected, is *not a proof* in the relevant sense, even if it is typeset as a proof and published in a mathematical journal. At best one can show that there are (fallacious) arguments, or proof-attempts, that have epistemic value. (We will do exactly this below.) And of course, it is unsurprising that fallacious arguments cannot be formalized as formal proofs in a consistent system. So what one would presumably need to do to make the argument go through would be to find an example of a proof that is widely accepted by the mathematical community as error-free, convincing, and otherwise laudable, but which is nonetheless not formalizable.[7]

We will not attempt to give an example of this sort. We do not think it is possible. Even if one offered a compelling candidate, we are skeptical that a convincing argument could be given that such a proof is not "formalizable". The reason is the essentially modal character of the demand: the question is whether there is *some* formal system or other such that the

---

[7] On this point, see de Toffoli (2024), who argues that what she calls "simil-proofs", which are apparent proofs that turn out later to have had errors, may nonetheless provide justification for believing results—and notes that very often such simil-proofs are treated as proofs in ordinary mathematical practice.

proof could be formalized as a (valid) proof in that system. This requirement is largely unconstrained. Moreover, given the nature of mathematical practice, and the standards for systematicity and rigor that mathematicians demand, one would expect that the arguments of any widely accepted proof could be made sufficiently explicit in some formal system or other—and if it turned out that existing formal systems were not up to the job, it would be an interesting project in mathematical logic to try to identify a new proof system in which the logic of this argument could be captured.

One might think that conceding this point is to concede that correctness is necessary after all. But we wish to point to three senses in which it is not. The first is simple: if we drop the strongly modal version of the requirement, then formalizability is clearly not necessary, by the very same reasoning. In other words: there is no formal system such that a given proof must be formalizable *in that system* in order for it to be epistemically valuable or otherwise mathematically acceptable. This claim is hardly new. At least on one reading (von Neumann, 1930/1983), Hilbert's program in foundations of mathematics was to show that any correct proof could be formalized in first order arithmetic (or, more generously, some fixed set theory)—which, in turn, he hoped, could itself be shown to be consistent *within that same system.* But of course, this goal foundered on the shoals of Gödel's incompleteness theorems, because any system in which Peano arithmetic (say) could be shown to be consistent would have to be strictly stronger than Peano arithmetic. Thus, when Gentzen (1964) later showed that Peano arithmetic is consistent, he used proof methods that could not be formalized in Peano arithmetic. And if one wished to show that Gentzen's system is consistent, one would need a stronger system still. And yet, the most committed advocates of formal correctness as a guiding norm in mathematics would surely want to concede that these sorts of arguments in the foundations of mathematics are epistemically valuable.[8]

---

[8]To be sure: we are not claiming that aiming to prove consistency results is deeply connected to the

One can see a different dynamic at play in the classic debate over intuitionism in the foundations of mathematics, though it leads to the same basic moral. For early intuitionists like Brouwer or Weyl, a proof could be epistemically valuable only if it was *constructive*, i.e., if existence proofs proceed by showing how to construct an instance of the things whose existence is claimed. This idea has led to several proposed systems of intuitionist logic, including the Brouwer-Heytin-Kolmogorov interpretation of constructive logic and Martin-Löf's intuitionistic type theory. Various authors have argued that mathematics must be formalizable in some such system in order to be epistemically valuable in the fullest sense, on the grounds that non-constructive proofs are unreliable. But such approaches have never gained widespread acceptance within mathematics, because intuitionistic approaches tend to be weaker than the methods of standard, informal mathematics. They limit what mathematicians can do. And when pressed, working mathematicians prefer to follow fruitful mathematical developments where they lead, rather than to constrain themselves to particular formal systems.

Towards the end of his life, reflecting on the debates in the foundations of mathematics on which he had cut his teeth from the perspective of someone who had shifted towards more applied questions, von Neumann summarized his attitude towards the earlier debates that we have been discussing as follows.

> In my own experience, on two other occasions in the early twentieth century, there were very serious substantive discussions as to what the fundamental principle of mathematics are; as to whether a large chapter of mathematics is really logically binding or not. And in the nineteen-tens and -twenties a critique of these questions made it apparent, that it was not at all clear exactly what one means by absolute rigor, and specifically, whether one should limit oneself to use only those parts of mathematics which nobody questioned. Thus, remarkably enough, in a large fraction of mathematics there actually existed differences of

---

Standard View. Consistency results in mathematical logic should be seen as mathematical theorems like any other. Our point is that you cannot fix a single formal system and insist that epistemic value is identified with formalizability in that system.

opinion! Some mathematicians said that one need not question any part of what is in fact being used. There was also a body of opinion, that one should not use more than what the most exacting critics had approved. However, there was a further, large body of mathematicians, who felt that while there was some point in questioning certain areas of mathematics, it was all right to use them. This group was quite ready to accept something like this: Those portions of mathematics which had been questioned and which had been clearly useful, specifically for the internal use of the fraternity—in other words, when very beautiful theories could be obtained in those areas—that those were after all at least as sounds as, and probably somewhat sounder than, the constructions of theoretical physics. And after all, theoretical physics was all right; so why shouldn't such an area, which had possibly even served theoretical physics even though it did not live up to 100 per cent of the mathematical idea of rigor, why shouldn't this be a legitimate area in mathematics; and why shouldn't it be pursued? This may sound off, as well as a bad debasement of standards, but it was believed in by a large group of people for whom I have some sympathy, for I'm one of them. (von Neumann, 1961, pp. 480-1)

When confronted with a choice between adhering to some fixed standard of rigor – of correctness – and valorous mathematical theories – beautiful ones, or deep ones, or fruitful ones – a significant portion of mathematicians will always prefer the latter. We contend they are right to do so—and for precisely this reason, formalizability in any particular system cannot possibly be deemed necessary for epistemic value in mathematics.

5. Fruitful Errors

In the previous section, we made the argument that formalizability in any particular system cannot be necessary for mathematics. We concluded with a passage from von Neumann pointing out that in cases where fruitful mathematics was found to run afoul of some formal system or other, mathematicians tended to reject the formal system and keep the fruitful mathematics. In this section, we wish to reflect further on a suggestion we find in that passage – taking, perhaps, some liberties of elaboration – to the effect that what ultimately grounds the epistemic value of some area of mathematics, or even some particular proof, is how that piece of mathematics fits into a broader network of mathematical and generally

scientific activities. We have used the term "fruitful" several times here, and what we have in mind is that fruitful mathematics is mathematics that leads to yet more of the same sort of activity, to new ideas and applications in physics or other mathematical sciences, or even just to yet more mathematics.[9] Mathematics that is not rigorous but which leads to successful applications can very likely be made rigorous, though doing so can itself be a challenging (and rich) task; but mathematics which is formally correct but not guided by fruitful application is not at all certain to ever find fruitful application—for reasons we have already argued for above.

This observation leads us to the second sense in which correctness is not necessary for mathematics, alluded to above: as a purely descriptive matter, within ordinary mathematical practice, correctness does not need to be established in order to secure the broad acceptance of a new result. Clearly, to publish a new proof in a top mathematics journal, it is not necessary to first formalize it. This, of course, is the observation that has driven much of the debate over the Standard View; and which has led Avigad (2021) to argue that the norms and strategies of informal proof can be seen as guides towards formalization without actually securing formalizability. But in fact, something deeper is true. Cases of mathematical *disagreement* are almost never resolved by formalizing of a difficult and controversial argument. When a major conjecture is settled using new techniques, for instance, and the mathematical community does not immediately accept it, one does not find mathematicians on each side of the dispute rush to produce formalizations or counter formalizations of the disputed proof. New techniques are not certified in this way.

What mathematicians *do* care about is ensuring that the new piece of mathematics is sound, in the sense that it preserves truth about whatever the target structure under consideration is. They want to be sure that counterexamples to the new result cannot be constructed using standard methods, that they will not inadvertently introduce contradic-

---

[9]Fruitfulness, here, is closely connected to what Maddy (2011) calls "depth" (see also Ernst et al., 2015).

tions by taking the result for granted, and that future applications of the theory will be reliable. In principle, formalization could perhaps help with this, insofar as one could show that some new result was equiconsistent with some widely accepted set of axioms (say) using certain generally reliable moves. But it not the most straightforward, perspicuous, or even reliable way of doing it. This is both because translating informal proofs into formal ones is not, in general, easy; it does not generally get to the heart of the subject matter; and *failures* to formalize in some particular system are not necessarily probative, for reasons already discussed. And so correctness is not necessary in the colloquial sense, either, since it is not required or needed to convince mathematicians of the reliability of a disputed result.

Finally, we turn to a third argument. We have thus far conceded above that "proof" is a success term, and that plausibly a proof that cannot be formalized, in any system at all, should not be deemed a proof, and we have argued that there are still senses in which formalizability is not necessary. But what if we forget about proofs, per se, and consider instead the more general category of proof candidates, including serious attempts at proof that may even appear in the published literature, but which turn out to be incorrect. There are many examples, we claim, of proof candidates that ultimately turned out to be incorrect—but which had epistemic value, as witnessed by their impact on subsequent mathematics. We will describe two such examples here.

In 1895, Poincaré released the first installment of his monumental paper *Analysis Situs*. His stated goal in this paper was to launch a new field of mathematics, now known as "algebraic topology". To say that he succeeded is to do scant justice: algebraic topology grew throughout the twentieth century and into the present to be one of the major fields of mathematical research, with influences and cross-pollination across mathematics, and across the sciences more generally. Today it finds significant applications in condensed matter physics, quantum chemistry and contemporary data science, just to name a few. *Analysis Situs* is also riddled with errors, which led Poincaré to produce five additional

"supplements". Even these did not suffice to fully clear up the issues, neither to today's sensibilities nor to those of his contemporaries.

It is hard to square the success and impact of these papers with the formalist view of mathematics. Standard formalist responses have alternated between treating them as anathema[10] and ignoring them altogether. Our goal here is not to rehash this discussion, which others, most recently McLarty (2024) have engaged with. Rather, we want to focus on the fecundity of Poincaré's mistakes, most notably his false assertion that every homology 3-sphere is homeomorphic to a 3-sphere, which led to his discovery of his dodecahedral space and the formulation of the "Poincaré conjecture".[11] This has been one of the most fertile mathematical conjectures of the last century and a half, leading to Smale's work on h-cobordism (and the proof of the conjecture in dimension ≥ 5) (Smale, 1962), Freedman's work on the disc-embedding theorem and the classification of simply connected (topological) 4-manifolds (Freedman, 1982), and Thurston's geometrization conjecture (Thurston, 1982)/Perelman's theorem classifying the (piece-wise) uniform geometries on closed 3-manifolds (Perelman, 2002, 2003b,a). For those keeping score at home, these results alone comprise three Fields Medals awarded, one Fields medal declined and one ($1,000,000) Millenium Prize declined (the only one which could have been awarded to date). Notably, Poincaré's errors are only the first in a series of major errors which have given rise to large swaths of contemporary algebraic topology and algebraic geometry, including:

1. Pontrjagin's famous incorrect computation of the stable homotopy groups of spheres in Pontrjagin (1938), leading to the "Pontrjagin-Thom" construction (Pontrjagin, 1947, 1955; Thom, 1954), Thom's work on cobordism theory (Thom, 1954), and the formulation in (Kervaire, 1960) of the "Kervaire invariant one" problem (solved by Hill et al.

---

[10]e.g. "For Bourbaki, Poincaré was the devil incarnate" (Mandelbrot, 2004, p. 280). The quote is also attributed to Stone (Mac Hale, 1993, p. 145).

[11]For a careful discussion, see the translator's introduction to Stillwell's translation of *Analysis Situs* (Poincaré, 2010).

(2016), except for one open case whose solution has recently been announced by Z. Xu).

2. Lefschetz's work on the topology of algebraic varieties, which led Hodge to initiate what we now call "Hodge theory" and the "Hodge Conjecture" (an open Millenium Prize problem). A classic quip, repeated by Griffiths in his biography of Lefschetz (Griffiths, 1992, p. 289), asserts that "Lefschetz never stated a false theorem nor gave a correct proof." Notably, one of the core theorems in question is a refinement of Poincaré duality, the site of a different major error in *Analysis Situs* and the one discussed by McLarty (2024).

And this is to name only two of the most consequential and far-reaching errors!

For another example, consider Frege's monumental effort to formalize mathematics (Frege, 1893, 1903), which is frequently credited as the origin of the current formalization frameworks. On the eve of publication of the final volume of his effort to derive all of mathematics from logic applied to a few self-evident axioms, Frege received a letter from Russell showing that the naive set theory underpinning Frege's project contained unavoidable contradictions. In an appendix written as the book was going to press, Frege described the experience as follows:

> Hardly anything more unfortunate can befall a scientific writer than to have one of the foundations of his edifice shaken after the work is finished. This was the position I was placed in by a letter of Mr. Bertrand Russell, just when the printing of this volume was nearing its completion.[12]

From one view, Russell's paradox mortally wounded Frege's project, with the final nail in the coffin delivered several decades later by Gödel's incompleteness theorems. However, as a historical matter, one could hardly hope for a more influential or productive mistake. The drama of Frege's "failure" provided a significant impetus and status-boost to research in

---

[12]Appendix of Frege (1903), in (Beaney, 1997, p. 279), translation by Michael Beaney.

axiomatic set theory and mathematical foundations, leading to several decades of work and attracting leading mathematicians including Zermelo, Hilbert, Ackerman, von Neumann, Gödel, Bernays, and Tarski, just to name a few. Even more, the error itself was and is a major discovery: every sufficiently expressive[13] finite formal system admits self-referentiality as an inescapable feature, and this radically limits the behavior of such systems. As a mathematical idea, this continues to bear fruit, from Gödel's second incompleteness theorems showing that no *finite* formal system is capable of satisfying Hilbert's criteria for an axiomatic framework for mathematics, to Turing's method for showing undecidability in theoretical computation, and then the many problems of mathematical and scientific interest for which no general algorithmic solution can exist.[14]

6. The Relationship Between Logic and Mathematics

Our discussion to this point has concerned "formal correctness", that is, correctness as determined by (valid) formalizability in some formal system. We have argued that formal correctness is neither necessary nor sufficient for epistemic value in mathematics. This raises the question of whether formal correctness bears any relationship to the epistemic value of mathematical proof at all. We suggest that the answer is "yes", but the situation is subtle. Our answer goes via a different notion of correctness, which might be called "mathematical correctness" (or "informal correctness"). "Mathematical correctness", here, is the standard of correctness for mathematical proof accepted by the mathematical community, on reflection. Somewhat more precisely, we take a lead from Peircean pragmatism and define mathematical correctness as "correctness at the end of inquiry", i.e., after indefinite scrutiny and in light of all subsequent mathematical developments.[15] Thus, results in early

---

[13] "sufficiently expressive"="sufficient to express integer arithmetic".

[14] At least in the Hilbert-Church-Turing paradigm of algorithm.

[15] For those worried about the apparent requirement of infinite time here, we suggest hearing this "weakly"; we are not aware of an instance where a span of more than a century of active research was required to detect an error in a central, widely credited result.

19th century analysis may be viewed as mathematically incorrect even though they were accepted by the community at the time. On the other hand, we do not want to suggest that the sociological whims of future mathematicians can retroactively dictate what makes good mathematics. One way to think about this worry is to say (somewhat speculatively) that Cauchy, Dirichlet, and their cohort would have accepted later arguments that aspects of their work were not mathematically correct, and in that sense the standards do not change.

Advocates for the Standard View would maintain that formal correctness is equivalent to mathematical correctness, or at least, necessary and sufficient for mathematical correctness. Indeed, there is a long tradition in philosophy of logic and mathematics, originating, arguably, with Leibniz's *characteristica universalis*, of conceiving of logic as providing a normative metaphysical and perhaps epistemic foundation for mathematics—or, in its most totalizing form, all human knowledge. This picture is one on which the rules for manipulating formal syntax are taken to be fundamental laws of reason, and various logical axioms are necessary truths that are prior to other forms of knowledge. From this perspective, logic is fundamental. Mathematics, meanwhile, is to be reduced to, or at least its truth is to be grounded in, logic. Arithmetic is true insofar as it can be reduced to a logical system; perhaps even our knowledge of arithmetic ultimately comes from its relationship to the secure truths of logic.

We propose a different – and we believe novel – view of the relationship between logic and mathematics. On our view, mathematical logic is best seen as a mathematical theory, or family of mathematical theories, of mathematical practice. It bears a relationship to ordinary mathematics analogous to the relationship that a mathematical theory in science bears to some real-world phenomenon. It captures some aspects of mathematics in precise mathematical terms. It allows us to prove theorems about what mathematics (so modeled) can and cannot accomplish, and it allows us to reason about the relationship between sentences, truth conditions, and structures that may instantiate what sentences assert. Among

other things, it provides a mathematical framework for thinking about how mathematicians reason, and it offers a precise account of mathematical correctness—namely, formal correctness.

From this point of view, mathematical logic is a valuable tool for thinking about mathematics. But to take formal correctness as the standard against which mathematical correctness is evaluated, or worse as equivalent to mathematical correctness, is to put the cart before the horse. It is analogous to saying that the reason the earth follows its orbit around the sun is that our best theory of gravitation says it ought to.[16] Or perhaps better, to say that a proof is mathematically incorrect because it is not formalizable would be like saying that Mercury has erred in some way because its perihelion advances a few dozen arcseconds each century relative to its Newtonian orbit.

We do not claim that the originators of mathematical logic thought of their work in the way we have just described. (Perhaps some did.[17] We are not sure. Others certainly did not.) Instead, we think of this view as a rational reconstruction, one that makes the best sense of the accomplishments of mathematical logic while also keeping track of its correct normative role in mathematical practice.

So why hold the view? The first reason is that mathematical logic clearly *is* a mathematical theory of mathematical practice. Classical first order logic captures many aspects of mathematical reasoning. First order model theory does describe a relationship between logical syntax and argumentation. Second order logic captures common reasoning in fields like

---

[16]Of course, it may well be that the central equations or laws of gravitational theory assert relationships that truly obtain in the world, and that those relationships are the reason that the earth follows its orbit around the sun. The point is that our theories attempt to describe the regularities that obtain in the world, rather than that the regularities in the world are somehow obligated to conform with our theories.

[17]For example, Gentzen (1964, p. 74) indicated that his goal was to develop "a formalism that reflects as accurately as possible the actual logical reasoning involved in mathematical proofs". Some philosophers read this "as meaning that all valid mathematical inferences should be instances of the logical rules of natural deduction, or closely related to them" (Tatton-Brown, 2023, p. 480), but we find that strained. The plain meaning of the text is that mathematical inference comes first, and the formalism of natural deduction was developed in its image.

point-set topology that typically use high order quantification. Intuitionistic logic captures the reasoning of mathematicians who reject certain proof methods allowed in first order classical logic. And so on. All of these are mathematical theories. So they are mathematical theories of mathematical practice. We infer from this that formal correctness, relative to a given system, is a precise way of capturing mathematical correctness within that system.

So far so good. But the important question is whether mathematical logic, and by extension formal correctness, is also *more* than this. In other words, one might argue that although mathematical logic is a mathematical theory of mathematical practice, it has a saliently different relationship to its subject matter than other mathematical theories of some natural phenomenon. We accept it does have some important differences, such as that it is self-referential. Insofar as mathematical logic is itself a part of mathematical practice, we can apply it to itself. But this is not a salient difference, because self-referentiality does not imply that mathematical logic has a normative status with regard to its subject matter, whereby failures of descriptive accuracy imply that the subject matter is erroneous.

One possible basis for a normative status of this kind would be if mathematical logic were well-conceived as an ideal form of mathematical reasoning, something like a perfection, precisification, or elaboration of ordinary mathematical practice. In that case, one might argue that mathematical logic is mathematical reasoning with certain failure modes removed, or that if a proof cannot be formalized, it fails some higher, more pure or ideal standard. But we would reject this. Yes, mathematical logic describes a certain kind of ideal of reasoning – or multiple ideals, associated with different logics – but that is because its a highly *idealized* theory of mathematical reasoning. By this, we mean that various bits of mathematical logic describe more or less simplified versions of mathematical reasoning, with domains of application in which they are more or less descriptively accurate, and other domains where they break down.

In fact, we have already seen this. First order logic has a number of nice properties

that capture deductive reasoning, it has a clear semantics, and one can prove fundamental theorems – of soundness and completeness – that link syntax and semantics. But standard methods in some areas of mathematics involve, and arguably require, second order quantification. This signals a breakdown of first order logic as an adequate model for mathematical practice. Of course, it does not mean that these methods cannot be formalized—one just needs to move to second order logic. But second order logic, while apparently expressive enough to capture classical mathematics, does not have the same nice semantic or proof theoretic properties as first order logic. It is famously a consequence of Gödel's incompleteness theorems that under standard semantics, there is no (finitary) proof system that satisfies all three of soundness, completeness, and effective proof checking. One can adopt alternative semantics (most notably, Henkin semantics), relative to which the standard deductive systems for first order logic become sound, complete, and effectively checkable, but Henkin semantics reinterprets second order quantification in a way that effectively reduces its expressive power, limiting its ability to faithfully model mathematical practice in the wild. So while second order logic improves on first order logic in some regards, it breaks down under other circumstances. Neither seems to perfectly capture mathematical practice, though both are helpful in certain domains.

Something similar can be said for mathematical arguments that move between proofs within a theory to proofs at the semantic or metamathematical level. As Rav (1999) points out, there are many examples in mathematics – he highlights group theory in particular – where a first-order theory is available, and for some purposes, textbook theorems can be seen as informal arguments that would admit formalization in a sound first-order deductive system. But often, powerful theorems of group theory require one to manipulate groups using the full resources of set theory, that is, to step outside of the first order theory of groups and prove things externally about structures that satisfy those axioms. Or consider that for standard results in synthetic differential geometry, one needs to move to an intuitionistic

logic and drop the law of the excluded middle. In that case, there are good reasons to think the results are mathematically correct, because the geometric theory is dual to an algebraic theory that does not require any changes in logic. But a flat-footed attempt to recover those results as formally correct leads to contradictions.

It is these limitations of mathematical logic, i.e., that no single deductive system can capture all aspects of mathematical practice, that we take to definitively show that formal correctness cannot be identical to mathematical correctness. Of course, none of these arguments imply that formal correctness, relative to some proof system, is not a valuable tool for assessing mathematical correctness. In many cases, showing that a proof is not formally correct, or could not be formally correct, relative to some system is probative in assessing whether it is mathematically correct, since the failure of formal correctness can reveal unnoticed errors. But we should think of this as a case of reasoning about some subject matter by using a simplified model that we expect to represent salient features for the case at hand. It is analogous to an engineer who uses mathematical models of material strain in the course of designing a bridge. The models are not the *reason* the bridge will stand or fall, and they can even be wrong under some circumstances. But they are likely a useful guide nonetheless, and will point to the places where the bridge design may be flawed.

We conclude this section with a final argument, which is somewhat speculative but which seems to us to accurately reflect both the sociology of mathematics and the relationship between logic and mathematics. Perhaps it is a Rorschach test of sorts. Suppose that there were a result that could not be formalized in any known logical system, but which nonetheless the mathematical community came to accept as mathematically correct. This may seem implausible—though we would argue the reason it seems implausible is that mathematical logic is a mature discipline, and there are many systems available already. But suppose it happens. What should we make of that situation? Advocates of the Standard View would have to say that the mathematical community has gone off the rails, they have accepted an

incorrect result, where incorrectness is determined by failures of formalizability. But that is not what we would expect mathematicians or logicians to actually do. Instead, we think such a situation would spark a lot of excitement among logicians, and important new work trying to find a *new* formal system in which the result could be formalized. One would try to understand how the result works and why it might be taken to be correct. In other words, if existing mathematical theories of mathematical practice failed in some domain, the correct and likely response would be to develop new mathematical theories of that domain.

7. The Automation of Mathematics Revisited

As we said in the introduction, one motivation for our arguments is to reflect on what automated theorem provers can contribute to mathematics—and to push back against the idea that automating mathematics is certain, or even likely, to significantly advance mathematics. We now return to this motivating theme.

As we noted in the introduction, if one thinks that formal correctness is the guiding virtue of mathematical practice, if one believes it is sufficient, or even just necessary, for epistemic value, then it would follow that a computational system that could reliably generate correct proofs would generate epistemic value. This would be all the more true if the computational system could reliably generate proofs faster than humans could generate them; or if it could generate proofs of theorems that humans have not been able to, or even in principle could not, prove. If, on the other hand, correctness is neither necessary nor sufficient for a proof to have epistemic value, it is much less clear what reliably generating correct proofs can accomplish. It would seem that mathematicians need more than this—and also, that providing correct proofs is not enough. Where does that leave us?

The first thing to say is that there is a sharp distinction to draw between computational proof generation and proof assistance. Using computers to help prove theorems that have been stated by mathematicians, which assert things that mathematicians have antecedent

28

interest in establishing, is surely valuable. Proof assistants can check potentially problematic lines of reasoning and provide some insight into their status; in principle, they can also fill gaps and play other kinds of supporting roles. None of this is problematic, as long as the crucial role that human mathematicians play in posing questions and extracting new understanding from answering them remains in focus. For that to be the case, the inputs – the conjectures stated, the definitions used, etc. – must come from mathematicians who are seeking to extend the body of existing mathematical knowledge, with full integration with the rest of that body of knowledge; and the outputs, namely, proofs of theorems, must be such that they can be read and understood by human mathematicians. Even more, these outputs must provide the kind of insight into the structures under consideration, and the reason the theorem is true, that other, non-assisted work does. In other words, proof assistants can surely play a collaborative role in producing new mathematical research, as long as the overall character of the resulting inquiry is not adversely affected.

More troubling, though, is the prospect that generative AI, such as high-end large language models, could *simulate* the sorts of inputs and output just described. That is, a system trained on mathematical definitions could surely state definitions that are similar in form to ones mathematicians would state; and they could pose conjectures with the right sort of form; and then automated theorem provers could generate proofs of them. One could even imagine (and there are ongoing efforts to build) AI systems that could take the output of a formal proof assistant / generator and rewrite it in the style of a working mathematician. In some ways this would be similar to generating prose in the style of various authors, or generating working computer code. One could even imagine that such a system could learn to accurately predict what sorts of definitions, conjectures, and theorems would be judged to be interesting by other mathematicians.

We say this is "troubling" because it could generate the sort of input and output that we described above as essential for preserving the role of human mathematicians (or something

indistinguishable from it), without humans having any role to play. In such cases, one might think that there is no loss in full automation, since the automated system would be generating "interesting" results in a form that is comprehensible to humans. But we think that even in this case, something crucial would be missing—though saying just what it is is more subtle. We note two kinds of problem (and DeDeo (2024) explores other, similar concerns). The first is that the sort of scenario we are imagining, if truly devoid of human feedback, would be static, whereas mathematics is dynamically evolving in response to new ideas, techniques, applications, and results. That is, what mathematicians today would call interesting is not what mathematicians fifty or a hundred years ago would call interesting—and if the field is healthy, fifty years from now, mathematicians' interests will have evolved further still. Making probabilistic judgments concerning what some fixed group of mathematicians would judge as interesting does not reflect that dynamics.

Of course, one could imagine a situation in which humans provide ongoing feedback to the AI mathematicians about what is interesting *now*, in light of changing background conditions. It might even be that mathematicians, responding to results generated by AI, would come to find new topics interesting, leading to new work by the AI, leading to new results, still—in each case, interesting within the evolving context. But we would argue that even in this case, something would be missing. This leads to the second problem, which is that the reasons the theorems are conjectured and proved would be the wrong reasons, and therefore they would not play the right role in the social activity of mathematics to count as contributions to that activity. The picture, here, is one of human discovery of structures and relationships, leading to new ideas about how to investigate those structures and what significance they have for other things of prior interest, leading to yet more ideas about those structures and new ones. The new ideas proceed from ever-deeper understanding of the old ones. This is an essentially creative activity, and a social one that stems from engagement with other mathematicians' work. The process by which it occurs is ultimately

what contributes most to our understanding of the mathematical subject matter, and, by extension to those parts of the world that we seek to describe with it.

No degree of simulation of this process via generation of the text and symbols that mathematicians produce to communicate with one another about mathematical ideas can play the cognitive role that reasoning about mathematics can play. This is true even if the simulated material is indistinguishable from the sorts of text that mathematicians themselves would produce. At very best, what one can hope for would be text that generates, in the mind of a mathematician who reads it, new interpretive and creative processes analogous to those that reading work by another mathematician would produce—thereby stimulating new ideas. But before one declares that this is enough, note that again the contribution to mathematics is what the mathematician produces when attempting to engage with the AI generated text, and not the text itself, which is essentially sterile.

Acknowledgments

References

Arnold, V., 2006. From Hilbert's superposition problem to dynamical systems, in: Bolibruch, A.A. *et al.* (Ed.), Mathematical Events of the Twentieth Century. Springer, pp. 19–47.

Avigad, J., 2006. Mathematical method and proof. Synthese 153, 106–159.

Avigad, J., 2008. Understanding proofs, in: Mancuso, P. (Ed.), The Philosophy of Mathematical Practice. Oxford University Press, Oxford, pp. 317–353.

Avigad, J., 2021. Reliability of mathematical inference. Synthese 198, 7277–7399.

Avigad, J., 2022. Varieties of mathematical understanding. Bulletin (New Series) of the American Mathematical Society 59, 99–117.

Azzouni, J., 2004. The derivation-indicator view of mathematical practice. Philosophia Mathematica 12, 81–105.

Beaney, M., 1997. The Frege Reader. Blackwell.

Bory, P., 2019. Deep new: The shifting narratives of artificial intelligence from deep blue to alphago. Convergence: The International Journal of Research into New Media Technologies 25, 627–642.

Bourbaki, N., 1970. Éléments de Mathématique, Théorie des Ensembles. Hermann, Paris.

Buzzard, K., 2022. The rise of formalism in mathematics. URL: https://www.youtube.com/watch?v=SEID4XYFN7o. iCM Plenary.

Buzzard, K., 2024. The fermat's last theorem project. URL: https://leanprover-community.github.io/blog/posts/FLT-announcement/.

Cellucci, C., 2008. Why proof? what is a proof?, in: Lupacchini, R., Corsi, G. (Eds.), Deduction, Computation, Experiment: Exploring the Effectiveness of Proof. Springer-Verlag, Milan, pp. 1–27.

DeDeo, S., 2024. Alephzero and mathematical experience. Bulletin of the AMS 61, 375–386.

Detlefsen, M., 2008. Proof: Its nature and significance, in: Gold, B., Simons, R.A. (Eds.), Proof and Other Dilemmas. Cambridge University Press, Cambridge, pp. 3–32.

Easwaran, K., 2009. Probabilistic proofs and transferability. Philosophia Mathematica 17, 341–362.

Ernst, M., Heis, J., Maddy, P., McNulty, M.B., Weatherall, J.O., 2015. Forward to special issue on mathematical depth. Philosophia Mathematica 23, 155–162.

Freedman, M., 1982. The topology of four-dimensional manifolds. J. Differential Geometry 17, 357–453.

Frege, G., 1893. Grundgesetze der Arithmetik, vol. I. Jena: Verlag Hermann Pohle.

Frege, G., 1903. Grundgesetze der Arithmetik, vol. II. Jena: Verlag Hermann Pohle.

Gentzen, G., 1964. Investigations into logical deduction. American Philosophical Quarterly 1.

Gowers, T., 2022. Announcing an automatic theorem proving project. Gowers Weblog URL: https://gowers.wordpress.com/2022/04/28/announcing-an-automatic-theorem-proving-project/.

Grayson, D., 2017. The mathematical work of vladimir voevodsky. URL: https://www.ias.edu/math/events/vladimir-voevodsky-memorial-conference-0. vladimir Voevodsky Memorial Conference.

Griffiths, P., 1992. Biographical Memoirs: Volume 61. The National Academies Press, Washington, DC. URL: https://nap.nationalacademies.org/catalog/2037/biographical-memoirs-volume-61, doi:10.17226/2037.

Hales, T., 2008. Formal proof. Notices of the AMS 55, 1370–1380.

Hamami, Y., 2018. Mathematical inference and logical inference. The Review of Symbolic Logic 11, 665–704.

Hamami, Y., 2022. Mathematical rigor and proof. Review of Symbolic Logic 15, 409–449.

Harris, M., 2019. Why the proof of Fermat's last theorem doesn't need to be enhanced. Quanta Magazine

URL: https://www.quantamagazine.org/why-the-proof-of-fermats-last-theorem-doesnt-need-to-be-enhanced-

Harris, M., 2022. The central dogma of mathematical formalism. URL: https://siliconreckoner.substack.com/p/the-central-dogma-of-mathematical.

Hersh, R., 1993. Proving is convincing and explaining. Educational Studies in Mathematics 24, 389–399.

Hersh, R., 1997a. Prove—once more and again. Philosophia Mathematica 5, 153–165.

Hersh, R., 1997b. What is Mathematics, Really? Oxford University Press, Oxford.

Hilbert, D., 1928. Die grundlagen der mathematik. Abhandlungen aus dem Seminar der Hamburgischen Universität 6, 65–85.

Hill, M., Hopkins, M., Ravenel, D., 2016. On the nonexistence of elements of Kervaire invariant one. Annals of Math. 184, 1–262.

Huntington, E., 1937. The method of postulates. Philosophy of Science 4, 482–495.

Kervaire, M., 1960. A manifold which does not admit any differentiable structure. Comm. Math. Helv 34, 257–270.

Lakatos, I., 1976. Proofs and Refutations. Cambridge University Press, Cambridge.

Larvor, B., 2012. How to think about informal proofs. Synthese 187, 715–730.

Larvor, B.P., 2022. On the unreasonable reliability of mathematical inference. Synthese 200, 332.

Mac Hale, D., 1993. Comic sections: the book of mathematical jokes, humour, wit and wisdom. Boole Press.

Mac Lane, S., 1986. Mathematics: Form and Function. Springer-Verlag, New York.

Maddy, P., 2011. Defending the Axioms. Oxford University Press, Oxford.

Mandelbrot, B., 2004. Fractals and Chaos - The Mandelbrot Set and Beyond. Springer.

Manin, Y.I., 1998. Interrelations between mathematics and physics, in: Séminaires et Congrès. volume 3, pp. 158–168.

McLarty, C., 2024. Poincaré on the value of reasoning machines. Bulletin of the AMS 61, 411–422.

von Neumann, J., 1930/1983. The formalist foundations of mathematics, in: Benacerraf, P., Putnam, H. (Eds.), Philosophy of Mathematics: Selected Readings. 2nd ed.. Cambridge University Press, Cambridge, pp. 61–65.

von Neumann, J., 1961. The role of mathematics in the sciences and in society, in: Taub, A.H. (Ed.), The Collected Works of John von Neumann. Pergamon Press. volume 6.

Perelman, G., 2002. The entropy formula for the Ricci flow and its geometric applications. arXiv:math/0211159 .

Perelman, G., 2003a. Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. arXiv:math/0307245 .

Perelman, G., 2003b. Ricci flow with surgery on three-manifolds. arXiv:math/0303109 .

Poincaré, H., 1905–1906. Les mathématiques et la logique. Revue de Métaphysique et de Morale, XIII:815–35 and XIV:17–34, 294–317 .

Poincaré, H., 2010. Papers on Topology: Analysis Situs and Its Five Supplements, trans. by John Stillwell. AMS and London Mathematical Society.

Polya, G., 1954. Mathematics and Plausible Reasoning. Princeton University Press, Princeton.

Pontrjagin, L., 1938. A classification of continuous transformations of a complex into a sphere. ii. C. R. (Dokl.) Acad. Sci. URSS , 361–363.

Pontrjagin, L., 1947. Characteristic cycles on differentiable manifolds. Mat. Sbornik N.S. , 233–284.

Pontrjagin, L., 1955. Smooth manifolds and their applications in homotopy theory. Trudy Mat. Inst. Steklov 45.

Rav, Y., 1999. Why do we prove theorems? Philosophia Mathematica 7, 5–41.

Scholze, P., 2020. Liquid tensor experiment. URL: https://xenaproject.wordpress.com/2020/12/05/liquid-tensor-exp

Smale, S., 1962. On the structure of manifolds. Amer. J. Math. 84, 387–399.

Tanswell, F., 2015. A problem with the dependence of informal proofs on formal proofs. Philosophia Mathematica , 295–310.

Tanswell, F.S., 2024. Mathematical Rigour and Informal Proof. Cambridge University Press, Cambridge, UK.

Tao, T., 2024. Embracing change and resetting expectations. Microsoft Unlocked - AI Anthology URL: https://unlocked.microsoft.com/ai-anthology/terence-tao/.

Tatton-Brown, O., 2023. Rigour and proof. The Review of Symbolic Logic 16, 480–508.

Thom, R., 1954. Quelques propriétés globales des variétés différentiables. Comm. Math. Helv 28, 17–86.

Thurston, W., 1982. Three dimensional manifolds, Kleinian groups and hyperbolic geometry. Bull. of the AMS 6, 357–381.

Thurston, W.P., 1994. Proof and progress in mathematics. Bulletin of the American Mathematical Society 30, 161–177.

de Toffoli, S., 2024. Proofs for a price: Tomorrow's ultra-rigorous mathematical culture. Bulletin of the

American Mathematical Society 61, 395–410.

Weir, A., 2016. Informal proof, formal proof, formalism. The Review of Symbolic Logic 9, 23–43.