



Reliability of mathematical inference

Jeremy Avigad^{1,2} 

Received: 28 August 2019 / Accepted: 17 December 2019 / Published online: 14 January 2020
© Springer Nature B.V. 2020

Abstract

Of all the demands that mathematics imposes on its practitioners, one of the most fundamental is that proofs ought to be correct. It has been common since the turn of the twentieth century to take correctness to be underwritten by the existence of formal derivations in a suitable axiomatic foundation, but then it is hard to see how this normative standard can be met, given the differences between informal proofs and formal derivations, and given the inherent fragility and complexity of the latter. This essay describes some of the ways that mathematical practice makes it possible to reliably and robustly meet the formal standard, preserving the standard normative account while doing justice to epistemically important features of informal mathematical justification.

Keywords Mathematical proof · Formalization · Reliability · Robustness

1 The standard view

Whatever the goals of mathematics are, surely one important goal is to get the answers right. Our calculations are supposed to be correct. Our mathematical claims are supposed to be justified by proofs, and those proofs are supposed to be correct as well.

Modern logic offers idealized accounts of what it means to be correct, based on the notion of a formal derivation. On that account, the mathematical objects we talk about are defined in a suitable formal foundation, our theorems are expressed as formal statements in that system, and our proofs are required to conform to the axioms and rules. Similarly, the expressions we calculate with denote objects that can be so defined, and our calculational steps are underwritten by theorems of the system. Mathematics is distinguished by the precise manner in which claims are stated and justified, and the logical account does it justice in that respect.

✉ Jeremy Avigad
avigad@cmu.edu

¹ Department of Philosophy, Carnegie Mellon University, Pittsburgh, USA

² Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, USA

Of course, ordinary mathematical texts are not nearly as regimented as formal proofs. But, on the logical story, they can be seen as an approximation to the ideal. Anyone putting forth an informal proof should be prepared to respond to queries for more information: “How is this term defined?” “Why does this claim follow?” Each response can engender further questions of the same sort, but, the thought goes, the process has to be bounded: if the proof is correct and the mathematician defending it sufficiently capable, the process cannot but terminate with terms whose meanings admit no further clarification, and with inferences whose correctness is part of the shared understanding. A formal logical foundation is designed to provide the fundamental terms and atomic rules of inference.

Two objections spring to mind. One is that a formal system cannot possibly be the operant standard because most mathematicians cannot even state the axioms of a particular formal system. The other is that there are various formal foundations on offer, and who is to say which is the right one? Even if we restrict attention to set theory, there are multiple versions and formulations in the literature.

Both objections can be met with a slight elaboration of the story above. When talking about formal foundations, we should keep in mind that the foundations of a building are typically below ground, and that we usually enter a building on the ground floor. We only need to descend to the basement when something goes wrong, and we are generally happy to leave it to workmen and contractors to look after the foundations while we tend to other things. So, too, when we deal with ordinary mathematics, we make use of a common language of numbers, sets, functions, relations, and structures. The objects we deal with can generally be described in these terms, built up from other objects that were previously defined in this way. Common inferences can be elaborated as needed, down to steps that are generally agreed upon as being part of the shared understanding. Logicians, when making the case that a certain formal system provides an adequate foundation for most mathematical arguments, show how to define the basic components in terms of the primitives of the system in question, and how to justify fundamental mathematical inferences. These serve to clarify mathematical language and shore up our proofs, but for the most part the details are not so important. The various foundational systems on offer are generally inter-interpretable, and they can all serve equally well.¹

One might object that not all foundational systems are equivalent: adding strong large cardinal axioms to set theory yields more theorems, and, in the other direction, intuitionistic logic imposes restrictions. But in this day and age, classical reasoning is commonly accepted in mathematics, and the vast majority of mathematics does not require strong axioms. And even in these extreme examples, mathematicians are generally aware that they are working constructively or making use of strong assumptions. So, even in these cases, we can still view mathematical arguments as being reducible to formal derivations, again with respect to a shared understanding of the methods and rules that are appropriate.

So, at least relative to a choice of axiomatic foundation, the view I am expounding here amounts to this. When someone in the mathematical community makes a mathe-

¹ Hamami (in press) does a particularly good job of clarifying the distinction between an informal foundation for mathematical reasoning and a formal one.

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.

This view has been held implicitly or explicitly throughout the philosophy of mathematics since the early twentieth century. Hamami (in press) finds clear expressions of the view in Mac Lane (1986) and Bourbaki (1966). Azzouni has more recently affirmed that the existence of a formal derivation is a contemporary standard of correctness:

The first point to observe is that formalized proofs have become the norms of mathematical practice. And that is to say: should it become clear that the implications (of assumptions to conclusion) of an informal proof cannot be replicated by a formal analogue, the status of that informal proof as a successful proof will be rejected. (Azzouni 2009, Sect. 4)

The viewpoint has also been articulated by Detlefsen (2008), who calls it the *common view*, Marfori (2010) and Hamami (in press), both of whom call it the *standard view*, and Burgess (2015, Chapter 2).

Recent developments in a branch of computer science known as *formal verification* provide empirical support, in that computational proof assistants have made it possible to construct fully formal computer-checked derivations of substantial mathematical arguments. Formal mathematical libraries now cover most subjects one comes across in an undergraduate mathematics education, including abstract algebra, number theory, analysis, topology, measure theory, and so on. Moreover, important mathematical theorems have been formalized, including the Feit–Thompson Odd Order Theorem (Gonthier 2013) and the Kepler Conjecture (Hales et al. 2017). (See Avigad (2018), Avigad and Harrison (2014) for surveys.) When someone working in the field embarks on a formalization project, the assumption that the theorem can be formalized is never in question, beyond perhaps the question as to whether the published proofs are correct. The question is, rather, how best to go about it and how long it will take. To be sure, the formalization process sometimes uncovers minor errors and omissions in the informal presentation that have to be remedied, but these are conventionally viewed as just that—errors and omissions—rather than indications that the informal source is correct but unformalizable.

But we do not need to invoke formal verification for empirical support. Mathematical disputes over the correctness of a proof generally hinge on the complexity of the argument and the difficulty of filling in the details, not on the correctness of any single, atomic inference. In practice, mathematicians generally expect that if a step in a proof is correct, it can be spelled out in detail, reducing it to inferences that are

mathematically immediate. And we have seen enough logic by now to know that such inferences are generally amenable to formalization.

The fact that some mathematical arguments draw on visual and spatial intuitions that are hard to spell out in symbolic terms is sometimes evoked to challenge claims of formalizability. But even though we still have a lot to learn about the nature of spatial and visual intuition in mathematics, the history and practice of mathematics only reinforce our central claim. It is well understood that some uses of diagrams in mathematical arguments can be straightforwardly (though painstakingly) eliminated. For example, diagram-chasing arguments in category theory can be spelled out in equational terms, even though there is generally no point to doing so. Other uses of visual intuition, for example in topological and geometric arguments, can be much harder to spell out in terms of textual inferences. But mathematics has taken great pains to develop the means to do that: we have definitions of spaces and manifolds and charts and atlases that are expressed in conventional set-theoretic terms, and these are designed to provide rigorous backing for our intuitions. Having such formal definitions is especially important to support reasoning about higher-dimensional spaces, where our intuitions are risky at best.

Indeed, diagrams play a very limited inferential role in contemporary mathematical journals. This is not to deny the importance of spatial and visual intuition, but it does indicate that we nonetheless expect that arguments based these intuitions can be formally justified. The nineteenth-century discovery of space-filling curves and nowhere differentiable functions only underscored the need to supplement intuitions with rigorous specification. The Jordan curve theorem, a historically salient example, asserts that any continuous loop in the plane that does not intersect itself divides the plane into two parts, an inside and an outside. The claim is intuitively clear, but modern mathematics has nonetheless seen fit to seek foundational justification. The Jordan curve theorem is in fact a theorem, and an important one. It has, moreover, been formalized (Hales 2007).

Despite all the things that can be said in its favor, however, the standard view has its opponents. Yehuda Rav, in particular, has criticized it on a number of grounds (Rav 1999), some of which have been addressed by Azzouni (2006, 2009). Others who have leveled criticisms or raised concerns include Marfori (2010), Detlefsen (2008), Tanswell (2015), Larvor (2016) and Pelc (2009). Some of these criticisms are discussed by Hamami (in press), who offers a spirited defense of the standard view.

My goal here is not to survey the arguments that have been marshalled on both sides of the debate. Rather, I will focus on one problem for the standard view, a problem which, I believe, lies at the heart of much of the criticism that has been leveled against it. In Sect. 2, I will lay out the problem and argue that we should take it seriously. The remainder of the paper is devoted to addressing it.

For all the reasons indicated above, I believe the standard view is essentially correct, and offers important insight as to what it means to acquire mathematical knowledge. But here I make the case that if we expect the view to weather the concerns that have been raised against it, we need to do more work to make it compelling, and develop a richer understanding of ordinary mathematical proof and its relationship to formal derivation. In Sects. 3 and 4, I take some initial steps toward that goal. Section 5 then situates this project with respect to broader issues in the epistemology of mathematics.

I argue there that even if we are primarily concerned with mathematical justification, adopting the standard view forces us to deal with a wider range of normative assessments, and to come to better terms with the nature of mathematical understanding.

2 The problem

According to the standard view, a mathematical statement is a theorem if and only if there is a formal derivation of that statement, or, more precisely, a suitable formal rendering thereof. When a mathematical referee certifies a mathematical result, then, whether or not the referee recognizes it, the correctness of the judgement stands or falls with the existence of such a formal derivation.

Our mathematical judgments are fallible, but, still, we hope that they are correct most of the time. Indeed, the status of mathematics as a rigorous discipline depends on that. The problem I would like to raise here is simply this: how can our mathematical judgments possibly warrant the existence of a formal derivation?

The problem is that formal derivations are very fragile objects, and are few and far between. The derivation of even a straightforward result in elementary number theory can require thousands of inferences in axiomatic set theory, and a sequence of a thousand inferences in which a single one is incorrect is simply not a proof. How can even the most carefully written journal article indicate the existence of a long sequence of inferences without a single error?

The probability of successful assessment decays exponentially with the length of the proof. Suppose there is a one percent chance of error in assessing the correctness of a step in a formal derivation generated by some fallible means. Then the odds of correctly assessing the validity of a proof of one hundred steps is about 37% (roughly $1/e$, where e is Euler's constant). At two hundred steps, the odds of success drop to less than 14%, and at four hundred steps, the odds drop to less than 2%.

Of course, informal proof texts are not formal derivations. On the standard view, they are only high-level sketches that are intended to *indicate* the existence of formal derivations. But providing less information only exacerbates the problem: if even a complete presentation of a formal derivation cannot be checked reliably, providing strictly less information can hardly provide more confidence. And the fact that ordinary mathematical proofs build on prior proofs in the literature makes matters even worse: any error in the chain that leads to the final result invalidates the conclusion. All things considered, on the standard view, it is hard to see how we can reliably do mathematics at all.² The fact that we obviously *can* reliably assess the correctness of at least some mathematical proofs makes it clear that there are mechanisms at play that are not accounted for by the standard view. We need to understand what those mechanisms are, and the question remains as to whether they have anything to do with formalization.

It seems to me that this general concern—the gap between informal proof and formal derivation, coupled with the inherent difficulty of making sense of and verifying the correctness of the latter—fuels much of the skepticism that has been expressed towards

² In his book, *Re-engineering Philosophy for Limited Beings* (Wimsatt 2007), William Wimsatt raises a similar concern about the robustness of long chains of reasoning in the context of scientific practice. See Chapter Four, titled “Robustness, Reliability, and Overdetermination.”

the standard view.³ It leads Rav to conclude that informal proofs do not function by indicating formal proofs, but, rather, by conveying semantic information:

Mathematical texts abound in terms such as “it follows from ... that,” “given that ... it is clear that” and the like; the antecedents are held to be true, from which the truth of the consequent is taken by *necessity* to follow. The issue is that beyond the *verbal* phrase “it follows from ... that so and so is the case” (equals “is true”), a mathematical proof in general only says *that* it follows, not *why* by logical necessity it has to follow. Hence the frequent need to interpolate further and further intermediate links as “bridges,” leading from claimed antecedents to the asserted conclusion. *Why* the consequent follows from the antecedents has to be figured out by the reader of a proof, based on the reader’s understanding of the *meaning* of the terms in the antecedent and consequent and requiring the reader’s familiarity with the underlying theory to which the proof is intended to be a contribution. None of these can be formally captured. (Rav 2007, pp. 316–317)

The mechanisms by which an informal proof conveys enough information for a competent referee to assess correctness are indeed complex, and I agree that this information is richer and more structured than a sequence of formal inferences. If Rav’s appeal to meaning, familiarity, and understanding helps convey the impression that there is something interesting going on, then I am all for it. But if the inquiry stops there, we have clearly failed. We are looking for a clarification of mathematical standards of correctness and insight as to how these standards are met, and vague appeals to meaning and understanding are at best optimistic gestures towards a more satisfying account.

Others, such as Detlefsen (2008) and Tanswell (2015), also dissociate epistemic justification of mathematics from its formalizability. For example, Detlefsen writes:

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. This notwithstanding, they are commonly presented in a way that *does* make their *rigor* clear—if not at the start, then at least by the time they’re widely circulated among peers and/or students. There are thus indications that rigor and formalization are independent concerns.

Although it is wise not to conflate informal rigor with formalizability, it is problematic to view them as fully independent. Severing ties completely requires us to provide an independent characterization of rigor and provide some other explanation of what it means for an informal proof to be correct. Formal logic was developed with precisely those goals in mind, and rejecting the formal account *tout court* relinquishes hard-earned gains without offering anything to replace them. Moreover, denying a correspondence between rigor and formalizability renders it utterly mysterious that

³ Pelc (2009) raises a more fundamental concern, namely, that complex theorems of mathematics may not have formal counterparts that can ever be verified mechanically by any processing device, given physical constraints such as the number of particles in the universe. But contemporary work in formal verification and interactive theorem proving establishes that they do, rather conclusively.

our mathematics *does* tend to be amenable to formalization, on those occasions when we set ourselves to it.

Azzouni, in contrast, maintains that informal proofs do in fact indicate the existence of formal derivations, though the nature of “indication” is subtle and complex (Azzouni 2004, 2006). But there is a tension in his writing: while defending formal systems as providers of public norms for correctness, he denies that they provide *epistemic* norms (Azzouni 2004, 2006, 2013). In other words, even though the existence of a formal derivation is the official standard, a mathematician need not be able to produce such a derivation in order to know or justify a claim. This raises the question as to how an informal proof serves as a reliable indicator. Azzouni only begins to address this, positing that we make use of higher-level *inference packages* that circumvent formal detail. According to Azzouni, such inference packages need not be formally describable (Azzouni 2013). More recent work suggests that such inference packages have an algorithmic component (Azzouni 2017).

Hamami is the most upbeat about the prospects of a formal account of the way we transform an informal proof to a formal derivation (Hamami in press). According to Hamami, we accumulate not just theorems but inferential patterns and procedures throughout our mathematical education, and we put them to good use towards understanding an informal proof. When a competent practitioner reads such a proof, the inferential cues and context trigger algorithmic processes that expand the inferential steps to smaller ones. These, in turn, are recognized as applications of prior theorems or lemmas, or at least patterns in memory that have been previously justified.

The accounts by Rav, Azzouni, and Hamami offer legitimate insights and, rightly, recognize the importance of background knowledge and expertise. But more is needed to convince the skeptic that any amount of background knowledge and expertise can be sufficient to bridge the gap between the informal and the formal. Even a short mathematical paper can include hundreds of inferential steps, and assessing each one involves a complex and generally fallible cognitive act. If an informal proof is really supposed to serve as a warrant for the existence of a formal derivation, there is little room for error in the checking process: acceptance of a single incorrect inference invalidates the entire judgment. Rav’s skepticism is therefore understandable. The standard view may be the best we have, but more work is needed to make it plausible.

Having set up the problem, let me close this section with two supplementary remarks. The first concerns the role that psychological or cognitive data should play in a philosophical study, and the second concerns the question as to whether the problem raised here is a philosophical problem at all.

I have framed the central problem in terms of a tension between the cognitive demands imposed by the standard philosophical view of mathematical proof and our inability to carry them out, given inherent constraints on our cognitive abilities. Moreover, the discussion has already invoked speculation about the cognitive processes that are involved in the evaluation of mathematical proofs. As a result, the problem may seem to belong more to psychology than philosophy. Indeed, I think it is useful to maintain a clear methodological distinction between the two subjects. Philosophy of mathematics aims provide an idealized account of mathematics, a normative practice for acquiring knowledge; the data to be explained is the body of mathematical knowledge, made manifest by the history of mathematics and its literature.

Psychology aims (among other things) to provide models of human cognition that are supported by observation and experimentation. Understanding our capacities for mathematical thought requires both. But my goal here is to explore what can be done on the philosophical side, by uncovering features of mathematics that serve to support reliable assessment. I will try to avoid detailed assumptions about human cognition, and simply recognize the fact that we have limited computational capacities, so that strategies that reduce the computational burden are generally advantageous. I hope that this exploration will support interaction between philosophical and experimental approaches to understanding the phenomena.⁴

The fact that the study I am engaged in here is not psychology does not, in and of itself, make it philosophy. From ancient times, philosophy has been concerned with identifying the proper grounds for mathematical knowledge and explaining what makes them so. This requires an idealization of mathematical practice, abstracting away those aspects of human activity that are only incidental in that respect. With that understanding, one might reasonably think that the problem raised here is merely a pragmatic question as to how to we can attain the philosophical ideal, rather than a part of philosophy of mathematics proper.

Constraining philosophy of mathematics in this way is a mistake. Our focus here remains on the proper means of acquiring mathematical knowledge, and general assumptions about our cognitive capacities have always been implicit in the subject. If we endow mathematical agents with a magical ability to grasp mathematical truths directly, there is not much for the philosophy of mathematics to do. Philosophical accounts have always presupposed that we have cognitive limitations, for example, that we are able to survey only finite amounts of data and carry out only certain sorts of inferential steps. The task of philosophy of mathematics is then to explain how we can have general knowledge of infinite domains—for example, how we can come to know things about all triangles or all numbers—given these constraints.

To make the problem here pressing, we only need to refine our assumptions slightly, and recognize that the resources we bring to bear on a mathematical task are not only finite, but also bounded. This means that we can only carry out tasks of limited complexity, and, generally speaking, inferential practices that reduce complexity have a bearing on our ability to acquire knowledge. In that light, the cognitive efficiency of our inferential practices is a legitimate normative goal. We can still idealize lots of things away while recognizing that an important feature of mathematics is its ability to manage complexity and thereby extend our cognitive reach. Our core concern is the same as that of Plato and Aristotle: understanding how we can, should, and do come to know mathematics.

3 General solutions

Our central problem rests on the fragility of formal proof, that is, the fact that an attempt at a formal derivation with even a single error is simply not a derivation. This is where the gap between informal proof and formal derivation is most keenly felt,

⁴ For an example of a fruitful interplay along these lines, see Hamami et al. (2019).

since an informal proof can have small mistakes and yet still reasonably lead us to believe in the correctness of its conclusion. It is not a contradiction in terms to credit a paper with establishing a certain result while at the same time noting that one of the lemmas in that paper is misstated, if the problem can be fixed and the fix is viewed as minor. If the standard view is correct, the informal proof still serves to indicate the existence of a formal derivation, without having to live up to the unreasonably high standard of being completely error free.

So the real question is how informal mathematical texts serve as reliable warrants to mathematical truth without spelling out every detail, while being *robust* with respect to error. Similar concerns for robustness arise in mechanical engineering, whose practitioners are required to develop systems and controllers whose behaviors remain within a suitable tolerance of the ideal, given variations in the design parameters, noise in the environment, defects in modeling and manufacturing, and so on. In a similar way, we need to understand how our mathematical practices can manage to reliably (though imperfectly) warrant the existence of formal derivations, given imperfections in their implementation and our abilities to process them.

If the first step towards a solution to our problem is to think of proofs, and the ambient mathematical environment, as artifacts that have been carefully designed to serve that purpose, the second step is to recognize that they, like other engineered artifacts, are highly structured objects. A proof is not just an undifferentiated sequence of inferential steps. For one thing, proofs are *modular*: definitions and lemmas serve to break arguments down into manageable components with well-defined interfaces (Avigad in press). Moreover, the presentation of an informal proof usually conveys a narrative structure. Morris has argued that the kinds of proofs we value tend to be *motivated*, which is to say, they convey information as to how individual steps are suggested by previous ones and how they constitute progress towards the final goal (Morris 2015, in press). Drawing on philosophical theories of planning and agency by Bratman (1999), Hamami and Morris have argued that the comprehensibility and effectiveness of an informal proof requires us to interpret them as instantiations of a *rational plan* (Hamami and Morris in press). With these insights, we can start to be more specific about the features of informal proof that convey understanding and support higher-level inference.

Taking these steps does not constitute progress in and of itself. Formal logic provides us a simple model of mathematical proof, and the normative requirements on assessment are straightforward: we simply have to match each step to the rules of a formal system. We have replaced this simple model with something more open-ended and vague, and we have not yet begun to explain how the new model delivers the reliability and robustness we are trying to explain. But, at least, it gives us more to work with.

It also opens the door to some encouraging metaphors. In engineering, modular design is commonly understood to result in more robust and maintainable systems. Comparing a proof to a narrative allows us to draw on intuitions regarding the distinction between the plot and its syntactic presentation. After reading *Pride and Prejudice*, we can reliably make claims about the characters and their motives, but we cannot reliably make claims about the number of occurrences of the letter “p.” If the judgment as to the correctness of a proof is more like the former, we have a chance; if it is more

like the latter, it is hopeless. Finally, thinking of proofs in terms of plans allows us to think about the strategies we use to maximize the likelihood of success. We might find it encouraging that we can generally assess the viability of a plan for traveling from New York to San Francisco, remaining robust with respect to small variations in the flight schedule, without knowing exactly how many pedestrians we will walk past on our way to the gate.

In the rest of this section, I will describe three general strategies for making mathematics robust, supported by this richer view of mathematical proof. At this stage, I will continue to rely on suggestive metaphor to get the ideas across, whereas, in the next section, I will discuss some mathematical strategies in more specific terms and relate them to the broad categories established here. The general strategies are as follows:

- Isolate and minimize critical information.
- Maximize exposure to error detection.
- Leverage redundancy.

There is a lot of data contained in a formal derivation. The first strategy involves minimizing the information that an informal proof must convey and that the receiving agent must process to ascertain that such a derivation exists. We can rely on the fact that a mathematical reader is expected to have suitable expertise, so we do not have to communicate every last detail. Rather, we can focus on the information that is sufficient to supplement the reader's knowledge and expectations.

Viewed in this way, an informal proof is a form of data compression. Coding schemes for data represent the most common patterns concisely, reserving extra bits for those that are unusual and unexpected. To invoke another analogy, software developers use version control software to store a project's history in a shared repository. To save space, the repository only has to store the difference between successive versions, rather than new copies. Similarly, video data can be transmitted more efficiently by sending only the differences between successive frames. The planning metaphor is also helpful: if I want to tell you how to travel from New York to San Francisco, I will assume you have a general familiarity with airports and taxis, and I will therefore focus on key details like the flight schedule and the address of the destination.

I will illustrate the second strategy with a story. I once heard the late Jon Borwein give a talk in which he displayed a calculation carried out by a computer algebra system. The calculation, a complicated sum, was glaringly incorrect: every term in the sum was clearly greater than or equal to zero, but the reported answer was negative. Borwein interpreted the brazenness of the error as a positive feature of the computer algebra system: when an incorrect answer looks right, we tend to accept it, but when an answer is obviously wrong, we know right away that there is more work to do.

Put simply, when a proof is wrong, we want it to be obviously wrong, or as clearly wrong as possible. We want our proofs to be open to vigorous probing, so that errors are easy to spot. This stands in analogy to Popper's dictum that good scientific theories should be open to falsification, though in this case the falsification comes from mathematical thought rather than from empirical data. Turning the etymological origins of the word "probability" into a slogan, we can describe the second strategy as that of maximizing the probability of correctness by maximizing probe-ability.

The two strategies described so far are complementary. The first focuses on the positive aspect of assessment, namely, corroborating the inferential steps that justify the conclusion. The second focuses on the negative aspect of assessment, namely, finding problems and spotting errors. The claim that both play a role in our understanding of proof is reinforced by research in automated reasoning, in which proof search can often be understood equally well as the search for a countermodel, whereby the failure to find a countermodel results in a proof, and vice-versa. (See, for example, proofs of the completeness of resolution (Bachmair and Ganzinger 2001, Sect. 3) and tableaux search (Hähle 2001, Sect. 3.4).) It also finds support from cognitive science. There is a longstanding debate as to whether we solve complex reasoning tasks by applying formal rules of inference or constructing mental models (Johnson-Laird 2010; Rips 1994), but most researchers are willing to acknowledge that both activities play a role.⁵ The claim that the assessment of a mathematical proof involves looking for inferential justification as well as looking for possible counterexamples is also well supported by the case studies and examples in Lakatos' *Proofs and Refutations* (Lakatos 1976).

The final strategy for increasing the reliability of our assessments is to leverage redundancy by ensuring that there are multiple ways to succeed. Our confidence in a plan to hike from one peak to the next is bolstered by the availability of multiple paths through the intervening ravine. Confidence in a plan to travel from New York to San Francisco is bolstered by having various means of getting to the airport, and knowing that there is a later flight in case the earlier one is canceled. The reliability of an engineered system is augmented by the existence of backup systems that can take over when a component fails, and bridges are made more reliable by the use of additional supports. A digital encoding of an image or text that renders the result incomprehensible if a single bit is lost is untenable, and the use of a redundant encoding can support error recovery.

This strategy is at odds with the first, since offering alternative means of proving a result requires conveying more information. But the payoff is often worth the cost. In this case, exponential decay works in our favor: if we can think of three ways of obtaining a desired result, each of which (independently) has a twenty-five percent chance of failure, the odds that they will all fail are only one in sixty-four, which is less than two percent.

An informal proof, if it is correct, can be formalized in lots of different ways. This means that insofar as an informal proof indicates the existence of a formal derivation, it indicates *lots* of them. Tanswell (2015) takes this underdetermination to pose a problem for the standard view, and he is right that any reasonable formulation of the view has to make room for it. On the approach I am urging here, the availability of multiple paths to formality is, on the contrary, a positive asset, and plays an essential role in the solution to our problem.

⁵ For example, even Philip Johnson-Laird, a foremost proponent of mental models, seems to acknowledge the use of inferential chains in solving logic problems (Johnson-Laird 2010, p. 5).

4 Specific solutions

It is time to leave the metaphors behind and turn to mathematics itself. My goal in this section is to consider some common features of mathematical practice and understand how they contribute to robust assessment along the lines described in the last section. I will cast these features in terms of a list of normative dictates, strategies that one might urge upon an aspiring young mathematician. In each case, I will argue that, first, these strategies *are* commonly employed in mathematics, and, second, that they do in fact contribute to robustness and reliability. The normative dictates are as follows:

1. Reason by analogy.
2. Modularize.
3. Generalize.
4. Use algebraic abstraction.
5. Collect examples.
6. Classify.
7. Develop complementary approaches.
9. Visualize.

I do not claim these strategies are exhaustive, and the discussion below will make it clear that there is substantial overlap. As a normative theory of mathematics, there is room for improvement. My goal is only to take a few steps in the right direction. Let us consider each item, in turn.

Reason by analogy If a proof I am trying to convey is similar to a proof you have seen before, then I only need to communicate the differences. For example, the fundamental theorem of arithmetic says that every nonzero integer can be factored into a product of prime numbers in a unique way, where “unique” means up to the order of the factors and up to multiplying by the units ± 1 . The same statement holds true of the Gaussian integers, which are the complex numbers of the form $a + bi$, where a and b are integers. (In this setting, $\pm i$ are included among the units.) Here, a proof of unique factorization can be obtained by drawing out analogies between the two domains. For example, given any pair of nonzero integers a and b , we can find a quotient, q , and remainder, r , so that $a = qb + r$, with $0 \leq r < |b|$. Almost the same fact holds true of the Gaussian integers; we only have to replace the last condition by $0 \leq \|r\| < \|b\|$, where $\|\cdot\|$ denotes the complex norm. This basic fact supports the transport of other facts from the regular integers to the Gaussian integers. If you are familiar with the proof of unique factorization for the integers, it is much easier to confirm that the result holds for the Gaussian integers as well. You only need to determine that the two domains are sufficiently similar for the proof to go through, and you can do that by focusing on the places where the differences may be problematic.

For another example, it is common in mathematics to embed a collection of objects in some sort of *completion*, a bigger space of objects with nicer properties. For example, the rational numbers can be completed to form the reals, and every partially ordered set can be embedded in its *ideal completion*. Sometimes such constructions can be viewed in more general terms; the construction of the reals from the rationals is an instance of a more general metric-space completion, and category theory often provides means of characterizing a completion in terms of a universal mapping property. But even in

the absence of such explicit generalizations, prior constructions make the new ones seem familiar and guide our reasoning about them. Arguments in functional analysis draw on analogies between spaces of functions and the simpler setting of finite vector spaces. Geometric operations on spaces, like projections and inner products, build on intuitions from high-school Euclidean geometry.

While the primary effect of reasoning by analogy may be to limit the information that needs to be conveyed, it also supports the second and third strategies. Casting an argument as analogous to a prior one enables the reader to focus attention on the differences and probe them to see if they break the proof. And reasoning by analogy helps preserve a multiplicity of approaches: any way of proving the prior theorem is a candidate for successful transfer.⁶

Modularize In a number of fields of engineering and design, a system or artifact is said to be *modular* to the extent that it can be decomposed into smaller components with limited interactions between them. These interactions are, in turn, mediated by well-defined interfaces. The notion of modularity, and the purported benefits of modular design, transfer well to mathematics (Avigad in press). The subject is inherently modular: one can use results from a distant branch of mathematics without knowing how they are proved, proofs are decomposed into smaller lemmas, and complex structures are defined in terms of simpler ones. In a detailed study of the history of Dirichlet's theorem on primes in an arithmetic progression, Morris and I were able to ascribe many of the benefits of later presentations to increases in modularity (Avigad and Morris 2016).

Modularity implements all three of the strategies enumerated in the last section. It serves to minimize the information that needs to be considered: anyone who knows the statement of a theorem can use it at once without reviewing the details of the proof, and anyone familiar with the fundamental properties of a complex object can make use of it without reviewing the details of its construction. Within a proof, keeping track of the flow of information imposes a cognitive burden, and breaking the proof into smaller components minimizes the information at play in each stage.

Modularity also supports error checking. Breaking the elements of a proof into smaller components makes it easier to check each one independently, much as is the case with modular hardware and software design. And modularity tends to limit the damage caused by an error, in that it is often possible to repair the proof of a single lemma without altering the rest of the argument. Finally, modularity supports redundant implementations: having well-defined interfaces makes it possible to swap out components for others with the same functionality.

Generalize Mathematical lemmas are often of the form $\forall x (A(x) \rightarrow B(x))$, where x is a sequence of objects, $A(x)$ represents a conjunction of some hypotheses, and $B(x)$ represents the conclusion of the lemma. In words, such a lemma says that any bunch of things satisfying A also satisfies B . The weaker the hypothesis, A , and the stronger the conclusion, B , the stronger the statement is, and hence the more likely to be wrong. In other words, the less we assume, and the more we conclude, the more forceful and risky the assertion. A conservative prover would therefore be induced to

⁶ For more on the use of analogical reasoning in mathematics, see Schlimm (2008, 2011).

state a lemma in the weakest terms needed to prove a theorem in question. But, in mathematical practice, we see just the opposite: the tendency is to state lemmas in the strongest terms possible.

This is one way that a mathematical proof exposes itself to error checking. If, in the middle of a complex proof with lots of objects and assumptions at play, I say “hence B ,” I am only asserting that B follows from everything that has come before. But breaking out the inference as a separate lemma and equipping it with the weakest hypotheses that are needed to make the conclusion hold serves a number of purposes. For one thing, it reduces the cognitive burden by making it easier to determine the relevant hypotheses. For another, it maximizes the exposure to error detection. Stating the weakest hypotheses possible dares the reader to find a counterexample. This is one common way that a proof turns out to be slightly wrong: an author may omit a hypothesis that is in fact available in the particular context she or he is interested in, and thus overgeneralize. Even in this case, the incorrect overgeneralization can help us confirm the correctness of the particular instance.

Generalization happens not only on the scale of individual inferences and lemmas, but also at the level of entire theories. In other words, it is sometimes possible to view an entire theoretical development as a generalization of a simpler one. Once the original theory is in place, the development of the more general one often relies on analogical reasoning, but now to a setting that subsumes the original one.

Once again, there is no shortage of examples. Probability theory deals with discrete spaces of events when analyzing dice, cards, and other games of chance. Extending the theory to continuous domains requires more work and more technical detail, but the definitions are designed to carry over intuitions from the discrete domain. Proofs are then similar, though they may require changing sums to integrals and carrying around more hypotheses. Understanding the discrete case makes it easier for us to check proofs in the continuous case, once again by focusing on the points where the differences threaten to cause problems.

Similarly, convex optimization is like linear optimization, but more general. In undergraduate calculus we learn that a continuous function on a closed interval attains its minimum and maximum values, and that if the function is moreover differentiable, the extrema occur either at the endpoints or at points where the derivative is zero. The field of convex analysis offers generalizations of these facts to multivariate functions on the real numbers, and then to infinite-dimensional analogues. Other generalizations treat functions that are not smooth everywhere, or multivalued functions. The more general results subsume the simpler cases and draw on the same intuitions. Familiarity with the simpler theory means that less information is needed to process the more complex one, and attention to the points of difference make it easier to locate potential errors.

Use algebraic abstraction The modern algebraic method involves characterizing classes of structures of interest by specifying the axioms they satisfy and then reasoning abstractly about these structures via this axiomatic interface (Schlimm 2011, 2013). The fact that the algebraic method incorporates all three previous strategies helps explain its importance.

To start with, the algebraic method is a prototypical means of generalization. When the group concept was first defined, particular instances had already proved their worth: groups of permutations, groups of geometric transformations, the additive group of integers modulo a fixed integer, and the corresponding multiplicative group of units (Wussing 1984). The development of an algebraic theory usually involves the transfer of intuitions from concrete cases, supporting the use of reasoning by analogy. And the algebraic method offers powerful forms of modularity: general definitions and theorems are isolated from the cluttered details of the specific instances, and axiomatic specification provides a clear interface to these instances.

Once a fundamental class of structures has been defined, the algebraic method continues to provide efficacious means of managing detail and transferring and reusing results. Both the integers and the Gaussian integers are instances of rings, and the quotient-remainder property they share amounts to the fact that they are both Euclidean domains. This, in turn, implies that they are principal ideal domains, and hence both unique factorization domains and Dedekind domains.

Other mathematical examples abound. Topological spaces, metric spaces, inner product spaces, and normed spaces all encapsulate intuitions about the Euclidean plane. This enables us to transfer and generalize arguments, for example, from finite to infinite dimensions. Other abstractions help us transfer intuitions about finite and discrete spaces to continuous ones. For example, compact topological spaces generalize finite spaces, and some of our intuitions about finite spaces apply to the compact setting.

Algebraic abstraction also provides us with means to probe for errors, by giving us a framework in which we can collect and organize potential counterexamples. The next two strategies indicate some of the ways it supports this.

Collect examples We have seen that it is often easier to prove a general theorem by proving a particular instance first and then transferring the more specific proof to the more general setting. Conversely, when assessing the correctness of a general proof, it is often helpful to think about what it says in a particular instance. After all, we may know enough about the particular instance to tell more easily whether intermediate statements in a proof hold in that setting. If they do, we are then in a better position to assess whether the reasons they hold carry over to the general setting. If they do not, we have detected an error in the argument.

This means that having an abundance of examples of structures and objects satisfying various properties is a virtue. And, indeed, we find that examples are highly valued in mathematics. Not only textbooks but also journal papers often follow a general definition with a number of examples. Mathematical readers have not only a stock of standard, typical cases in mind when they read a theorem, but usually also a battery of atypical ones, which serve as useful test cases. To support the latter, in addition to illustrative examples, textbooks and journal papers also often contain illustrative *counterexamples* to definitions and sets of hypotheses. Books like *Counterexamples in Topology* (Steen and Seebach 1970) and *Counterexamples in Analysis* (Gelbaum and Olmsted 1964) are even dedicated to that purpose.

Classify It is helpful not only to have plenty of examples at one's disposal, but, moreover, to have them categorized and sorted in useful ways. Groups can be finite, abelian, nilpotent, solvable, discrete, and so on. An expert mathematical reader can be expected

to know the lay of the land, and to recognize situations where various hypotheses are useful. The absence of an expected assumption may raise concerns. Producing an explicit example where the hypothesis fails yields a test case for the conclusion of an inference or lemma.

Suppose a group-theoretic argument uses the fact that $(ab)^n = a^n b^n$, a fact familiar from calculations in arithmetic. A competent mathematical reader will spot right away that the claim does not hold in general. The obvious proof strategy requires the group operation to be commutative, and knowing that there are nonabelian groups raises a red flag. In fact, the smallest nonabelian group, the dihedral group of order six, provides a counterexample.

Often a classification theorem in mathematics takes the form of a *structure theorem*, which represents members of a family of structures uniformly in terms of a set of parameters that characterize each structure up to isomorphism. For example, vector spaces are characterized by the cardinality of a basis, and every vector space can be seen as the set of finite linear combinations of basis elements. Every abelian group is a direct sum of cyclic groups, and this fact can be used to parameterize the collection. Having such canonical representations means that one can often prove a theorem in abstract terms or in terms of the explicit representations, providing complementary approaches. Classification also suggests parameters that can be varied in the search for a counterexample: is the theorem true of infinite-dimensional spaces? Does the proof presuppose that the basis is countable?

Develop complementary approaches Leveraging redundancy means fostering multiple ways of carrying out inferences, and one finds many instances of this in the mathematical literature. It is common to find different proofs of the same theorem; for example, there are more than two hundred published proofs of the law of quadratic reciprocity (Lemmermeyer 2000). This is not surprising, given that we learn different things from different proofs (Dawson 2015). But the phenomenon is relevant to our present concerns because it provides a means of triangulating a theorem and confirming its claims from multiple perspectives.

Mathematical presentations often foster triangulation on a smaller scale. Even in journal publications, it is common to see statements like “ X follows from Theorem Y in paper Z , but, for completeness, we provide a more direct proof here,” or “one way to prove this theorem would be to . . . , but we will follow a different approach here.” Even individual inferences are often given alternative justification. Such gestures towards alternative approaches lend confidence in the correctness of the result.

The algebraic method supports multiple perspectives by giving us ready-made ways of honing in on different aspects of mathematical structure. Given a mathematical structure, one can look for informative ways of viewing it as a topological space or a metric space; one can consider the lattice of substructures or the collection of homomorphic images; or one can situate it in a category of like objects. This often makes it possible to account for the validity of an inference in complementary ways.

Visualize Examples of visual intuition in mathematics are compelling, and of all the strategies described here, this one has received the most philosophical attention in recent years (De Toffoli 2017; De Toffoli and Giardino 2014, 2015, 2016; Giaquinto 2007; Hamami and Mumma 2013; Hamami et al. 2019; Mumma 2012; Tappenden

2005; Waszek 2018). Jill Larkin and Herbert Simon famously argued that diagrammatic representations leverage our spatial cognitive capacities by using features and relationships in the diagram to make relevant information salient (Larkin and Simon 1987). In this way, diagrammatic reasoning supports both efficient search for positive justification and probing for counterexamples. Features in a diagram can suggest properties that follow from the constraints that the diagram is supposed to satisfy. Features absent from a diagram, or that can be eliminated by varying the diagram in a way that preserves the constraints, give rise to counterexamples. Various authors have explored ways that this plays out in mathematical reasoning, including geometry, topology, and algebra (Avigad et al. 2009; De Toffoli 2017; De Toffoli and Giardino 2014, 2015, 2016; Giaquinto 2007; Manders 2008; Mumma 2012). In the case of Euclidean geometry, Hamami et al. (2019) provide experimental evidence to support such a cognitive model.

Visual intuition can even play a role in understanding proofs that make no use of diagrams. Mathematical terminology often involves spatial metaphors, and terms like “space,” “continuous transformation,” “distance,” “interior,” “endpoint,” and “smooth curve” can’t avoid invoking images from introductory textbooks and lectures. Even relations like ordering relations on the real numbers and set-theoretic containment relations are conventionally illustrated with diagrams, and we can call upon these images when reading any mathematical text. Modern mathematics provides ways of defining all these notions and reasoning about them in set-theoretic terms, but the absence of visual intuition from these representations is illusory, and it is likely that we use the same spatial intuitions to assess the correctness of inferences that involve these notions. If that is the case, then visual representation can be seen as providing efficient encodings that make it possible to probe and confirm text-based inference as well.

Summary Let me close this section by summarizing the way these specific strategies instantiate the more general ones enumerated in the last section. To isolate and minimize critical information in a proof, we can present an argument as a modification or generalization of something familiar, and look for representations that make the most important information salient. We can treat general results in the literature as black boxes, and focus on assessing their applicability. Within a proof, we can decrease the cognitive burden by modularizing and isolating the details that are relevant to each inference. We can maximize exposure to error detection by generalizing inferences and stating the implications in the strongest terms possible. The availability of examples and classificatory schemes provides means of probing inferences and testing their validity. Visual imagery, diagrammatic reasoning, and algebraic abstraction all help us organize and track information, as well as suggest counterexamples. Finally, the availability of multiple representations, and multiple routes to a conclusion, adds to the robustness of a proof.

5 Conclusions

The considerations in the last section point the way to augmenting Hamami’s model of rigor so that it can better address some of Azzouni’s and Rav’s concerns. I believe that

Hamami's model is essentially correct: when we read an informal mathematical proof, we really do try to expand inferences in order to gain confidence that a much more detailed version could be given, down to the kinds of basic inferences that twentieth century logic has shown can be reduced to axiomatic primitives. At the same time, we can be convinced by an informal proof without carrying out a fully detailed expansion, and it is too much to ask that we reach the point where each inference is an instance of a known theorem or an explicit rule we have stored in memory. Rather, we rely on something like Azzouni's inference packages to take bigger steps. Maybe we convince ourselves that the argument that needs to be filled in is sufficiently analogous to another argument we know. Maybe we confirm that a general conclusion holds in one particular case, and convince ourselves that the particular case is sufficiently representative of the general one. Maybe we have a visual argument in mind ("if we move this configuration to point a and the distances are small enough, then point b clearly has to lie inside the region in question") with confidence that the usual ways of spelling things out in topological, metric, or geometric terms provide a clear route to a more formal derivation. Because these judgments rely on heuristics and metacognitive reflection, Azzouni is right to think that it can be hard or impossible to describe them formally. And when they rely on domain-specific background knowledge, Rav is right to feel that they depend on the meaning of the terms involved and the reader's familiarity and understanding.

But this does not mean we that we cannot subject these kinds of inferences to careful study. The challenge is to understand how the strategies they incorporate provide means of acquiring knowledge that are reliable and robust, albeit fallible. Cognitive scientists may be able to fill out the picture by studying expert behavior in controlled environments. But philosophers of mathematics can also play a part, by paying close attention to ordinary mathematical proofs and the ways they are presented. The mathematical literature gives us a lot to work with, providing an inexhaustible source data for when it comes to understanding conventional mathematical inference.

This approach preserves the standard view that informal proofs work by indicating formal derivations, and it supports Hamami's claim that they do so relative to a reader with sufficient procedural capacities to process the informal proof and generate more detailed justification. The approach also augments the normative component of the standard view, since it emphasizes that in addition to being correct, a proof should support robust and reliable assessment. In mathematics, it is common to complain that a proof, though correct, is hard to read and understand. It is equally common to praise another proof for being clear, insightful, and to the point. These are clearly normative judgments, and the philosophy of mathematics can help us make sense of them.

I have surmised that much of the skepticism that has been directed towards the standard account stems from the obvious gap between informal proof and formal derivation. Philosophers of mathematics should take this skepticism seriously, since the ability of mathematics to bridge the gap is one of the most important and interesting aspects of the practice. Mathematics requires us not only to be correct, but to be correct about complex things. It requires us to solve hard problems and carry out long and subtle chains of reasoning in reliable and robust ways. The standard view therefore offers not just a standard of correctness but a greater philosophical challenge, and

understanding the sense in which the view is correct can tell us important things about how mathematics manages complexity.

At the same time, some of the skepticism that has been directed towards the standard view does not seem to target the view itself, but, rather, its collateral effects. Rav has emphasized that we get a lot more from a proof than a certificate of correctness, and that we value proofs for reasons that go beyond the ability to construct a formal derivation (Rav 1999). He may therefore feel that a narrow focus on formal correctness has distracted us from more interesting and important topics. Similarly, the criticisms one finds in Marfori (2010) and Tanswell (2015) challenge the standard view's ability to provide an adequate account of mathematical knowledge more than its ability to provide a normative standard of correctness. They and other authors have therefore begun to explore the broader epistemic benefits of informal proof (Avigad 2006, 2008; Larvor 2012; Robinson 2000; Tanswell 2016). It is even reasonable to maintain that philosophy of mathematics has suffered from excessive fixation on proof itself, and a number of philosophers of mathematics are turning their attention to aspects of mathematical understanding that extend well beyond justification and correctness (Avigad 2003; De Toffoli 2017; De Toffoli and Giardino 2014, 2015, 2016; Ferreiros 2016; Hamami and Morris in press; Hamami and Mumma 2013; Hamami et al. 2019; Mancosu 2008; Morris 2015; Mumma 2012; Schlimm 2008, 2011, 2013; Waszek 2018).

The point I wish to make here is that the standard view is fully compatible with these attempts to grapple with other aspects of mathematical understanding. In fact, the approach I have taken here provides a much stronger basis for reconciliation: I have argued that even if one is primarily concerned with standards of correctness, one *has* to come to terms with the mechanisms that support more general mathematical understanding in order to make sense of how mathematics makes it possible to meet those standards.

It is not a coincidence that many of the strategies I have enumerated play a more general epistemological role. Abstraction, generalization, modular structuring of mathematical theories and proofs, systematic organization and reorganization of our mathematical artifacts, and the development of multiple perspectives support all aspects of mathematical reasoning, not just our assessments of correctness. The recent philosophical work on visual and diagrammatic reasoning is similarly concerned not only with the roles that such reasoning plays by allowing us to “see” the correctness of an inference, but the also with the roles it plays in mathematical discovery and exploration.

These observations can help us interpret historical and contemporary debates in a new light. The nineteenth-century revolution in mathematics is often viewed as a battle between abstract, conceptual reasoning, embodied by mathematicians such as Dedekind, Riemann, and Cantor, against calculational reasoning, represented by the members of the Berlin school like Weierstrass and Kronecker. This, in turn, is often cast as a battle between the competing virtues of intuition and rigor. The famous Jaffe–Quinn debate (Atiyah et al. 1994) of the 1990s is often viewed in these terms. Here I have argued that concerns about intuition and concerns about rigor are not so cleanly separated. If we take deductive reasoning to be representative of calculational justification and the strategies I have enumerated as among the components a conceptual

understanding, my case is that both calculational and conceptual means are needed to support reliability and robustness.

Earlier in nineteenth century, in his work on solvability of polynomials, Galois raised concerns about the complexity of calculation:

If you now give me an equation that you have chosen at your pleasure, and if you want to know if it is or is not solvable by radicals, I could do no more than to indicate to you the means of answering your question, without wanting to give myself or anyone else the task of doing it. In a word, the calculations are impracticable.

From that, it would seem that there is no fruit to derive from the solution that we propose. Indeed, it would be so if the question usually arose from this point of view. But, most of the time, in the applications of the Algebraic Analysis, one is led to equations of which one knows beforehand all the properties: properties by means of which it will always be easy to answer the question by the rules we are going to explain. ...All that makes this theory beautiful and at the same time difficult, is that one has always to indicate the course of analysis and to foresee its results without ever being able to perform [the calculations]. [Galois (1962), pp. 39–40, as quoted and translated in Tignol (2016), Chapter 14]

One finds similar language in Dedekind, who, in developing the theory of algebraic ideals, aspired

...to seek proofs based immediately on fundamental characteristics, rather than on calculation, and indeed to construct the theory in such a way that it is able to predict the results of calculation...(Dedekind 1996, p. 102)

The desire to have means of seeing the results of a calculation without having to carry it out has been central to modern mathematics since its inception. It is informative to consider, in this light, the standard view's requirement that we establish the existence of a formal derivation without spelling it out precisely.

Once we situate concerns about the correctness of a mathematical proof in a proper context, the distinction between rigor and intuition becomes less sharp, and both the logic of discovery and the logic of justification come under a bigger umbrella (Feferman 1978; Lakatos 1976). Coming to terms with the nature of mathematical justification is not at odds with understanding a wider range of mathematical values, but an integral part of the greater enterprise.

To sum up, I have defended two complementary claims. The first is that the gap between informal proof and formal derivation is not a good reason to reject the latter as a normative standard of correctness for mathematical proof. The second is that accepting this standard is not at odds with the task of making sense of higher-level epistemic features of mathematical reasoning; rather, the latter is an essential prerequisite to understanding how the normative standard can and should be met. I have taken initial steps towards developing a positive account of some of the relevant epistemic features, and together we have begun to understand how they make reliable and robust assessment possible.

Supporters of the standard view may interpret the first claim as vindication, while its opponents may take the second claim to corroborate the inadequacy of a narrow focus on formality. But accepting both claims goes a long way to blunt the debate, and leaves us with a clearer understanding of what the standard approach does, and does not, accomplish. Perhaps, then, it is time to move on, and focus our attention on the important questions that remain.

Acknowledgements I am grateful to Silvia De Toffoli, Yacin Hamami, and three anonymous reviewers for helpful comments and suggestions. This work has been partially supported by AFOSR Grant FA9550-18-1-0120 and the Sloan Foundation.

References

- Atiyah, M., et al. (1994). Responses to: A. Jaffe and F. Quinn, “Theoretical mathematics: Toward a cultural synthesis of mathematics and theoretical physics”. *Bulletin of the AMS*, 30(2), 178–207.
- Avigad, J. (2003). Number theory and elementary arithmetic. *Philosophia Mathematica*, 11, 257–284.
- Avigad, J. (2006). Mathematical method and proof. *Synthese*, 153(1), 105–159.
- Avigad, J. (2008). Understanding proofs. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 317–353). Oxford: Oxford University Press.
- Avigad, J. (in press). Modularity in mathematics. *Review of Symbolic Logic*.
- Avigad, J. (2018). The mechanization of mathematics. *Notices of the American Mathematical Society*, 65(6), 681–690.
- Avigad, J., Dean, E., & Mumma, J. (2009). A formal system for Euclid’s Elements. *Review of Symbolic Logic*, 2, 700–768.
- Avigad, J., & Harrison, J. (2014). Formally verified mathematics. *Communications of the ACM*, 57(4), 66–75.
- Avigad, J., & Morris, R. (2016). Character and object. *Review of Symbolic Logic*, 9, 480–510.
- Azzouni, J. (2004). The derivation-indicator view of mathematical practice. *Philosophia Mathematica*, 12(2), 81–105.
- Azzouni, J. (2006). *Tracking reason*. Oxford: Oxford University Press.
- Azzouni, J. (2009). Why do informal proofs conform to formal norms? *Foundations of Science*, 14(1–2), 9–26.
- Azzouni, J. (2013). The relationship of derivations in artificial languages to ordinary rigorous mathematical proof. *Philosophia Mathematica*, 21(2), 247–254.
- Azzouni, J. (2017). Does reason evolve? (Does the reasoning in mathematics evolve?). In B. Sriraman (Ed.), *Humanizing mathematics and its philosophy: Essays celebrating the 90th birthday of Reuben Hersh* (pp. 253–289). Cham: Springer.
- Bachmair, L., & Ganzinger, H. (2001). Resolution theorem proving. In J. A. Robinson & A. Voronkov (Eds.), *Handbook of automated reasoning (in 2 volumes)* (pp. 19–99). Amsterdam: Elsevier.
- Bourbaki, N. (1966). *Éléments de mathématique*. Fasc. XVII. Livre I: Théorie des ensembles. Hermann, Paris.
- Bratman, M. E. (1999). *Intention, plans, and practical reason*. Cambridge: Cambridge University Press.
- Burgess, J. P. (2015). *Rigor and structure*. Oxford: Oxford University Press.
- Dawson, J. W, Jr. (2015). *Why prove it again?*. Cham: Springer.
- De Toffoli, S. (2017). ‘Chasing’ the diagram—the use of visualizations in algebraic reasoning. *The Review of Symbolic Logic*, 10(1), 158–186.
- De Toffoli, S., & Giardino, V. (2014). Forms and roles of diagrams in knot theory. *Erkenntnis*, 79(4), 829–842.
- De Toffoli, S., & Giardino, V. (2015). An inquiry into the practice of proving in low-dimensional topology. In G. Lolli, G. Venturi, M. Panza (Eds.), *From logic to practice* (pp. 315–336). Cham: Springer.
- De Toffoli, S., & Giardino, V. (2016). Envisioning transformations—the practice of topology. In B. Larvor (Ed.), *Mathematical cultures* (pp. 25–50). Cham: Birkhäuser/Springer.
- Dedekind, R. (1996). *Theory of algebraic integers*. Cambridge: Cambridge University Press. A translation of *Sur la théorie des nombres Entiers Algébrique* (1877), translated and introduced by John Stillwell.

- Detlefsen, M. (2008). Proof: Its nature and significance. In B. Gold & R. A. Simons (Eds.), *Proof and other dilemmas: Mathematics and philosophy* (pp. 3–32). Washington, D.C.: Mathematical Association of America.
- Feferman, S. (1978). The logic of mathematical discovery vs. the logical structure of mathematics. In *Philosophy of science association*, East Lansing. Reprinted in *In the light of logic*, Oxford University Press, 1998 (pp. 77–93).
- Ferreiros, J. (2016). *Mathematical knowledge and the interplay of practices*. Princeton: Princeton University Press.
- Galois, E. (1962). *Écrits et mémoires mathématiques*. Gauthier-Villars, Paris. Edited by Robert Bourgne and Jean-Pierre Azra.
- Gelbaum, B. R., & Olmsted, J. M. H. (1964). *Counterexamples in analysis*. San Francisco: Holden-Day.
- Giaquinto, M. (2007). *Visual thinking in mathematics: An epistemological study*. Oxford: Oxford University Press.
- Gonthier, G., Asperti, A., Avigad, J., Bertot, Y., Cohen, C., Garillot, F., et al. (2013). A machine-checked proof of the odd order theorem. In S. Blazy, C. Paulin-Mohring, & D. Pichardie (Eds.), *Interactive theorem proving* (pp. 163–179). Berlin: Springer.
- Hähle, R. (2001). Tableaux and related methods. In J. A. Robinson & A. Voronkov (Eds.), *Handbook of automated reasoning* (pp. 100–178). Amsterdam: Elsevier.
- Hales, T. (2007). The Jordan curve theorem, formally and informally. *American Mathematical Monthly*, 114(10), 882–894.
- Hales, T., Adams, M., Bauer, G., Dang, T. D., Harrison, J., Hoang, L. T., et al. (2017). A formal proof of the Kepler conjecture. *Forum of Mathematics, Pi*, 5, e2.
- Hamami, Y. (in press). Mathematical rigor and proof. *Review of symbolic logic*.
- Hamami, Y., & Morris, R. (in press). Plans and planning in mathematical proofs. *Review of Symbolic Logic*.
- Hamami, Y., Mumma, J., & Amalric, M. (2019). Euclidean diagrammatic reasoning as counterexample reasoning. Manuscript.
- Hamami, Y., & Mumma, J. (2013). Prolegomena to a cognitive investigation of Euclidean diagrammatic reasoning. *Journal of Logic, Language and Information*, 22(4), 421–448.
- Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 107(43), 18243–18250.
- Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. Cambridge: Cambridge University Press.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65–100.
- Larvor, B. (2012). How to think about informal proofs. *Synthese*, 187(2), 715–730.
- Larvor, B. (2016). Why the naïve derivation recipe model cannot explain how mathematicians' proofs secure mathematical knowledge. *Philosophia Mathematica*, 24(3), 401–404.
- Lemmermeyer, F. (2000). *Reciprocity laws: From Euler to Eisenstein*. Berlin: Springer.
- Mac Lane, S. (1986). *Mathematics. Form and function*. New York: Springer.
- Mancosu, P. (Ed.). (2008). *The philosophy of mathematical practice*. Oxford: Oxford University Press.
- Manders, K. (2008). The Euclidean diagram. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 80–133). Oxford: Oxford University Press.
- Marfori, M. A. (2010). Informal proofs and mathematical rigour. *Studia Logica*, 96(2), 261–272.
- Morris, R. (in press). Motivated proofs: What they are, why they matter, and how to write them. *Review of Symbolic Logic*
- Morris, R. (2015). Appropriate steps: A theory of motivated proofs. Ph.D. thesis, Carnegie Mellon University
- Mumma, J. (2012). Constructive geometrical reasoning and diagrams. *Synthese*, 186(1), 103–119.
- Pelc, A. (2009). Why do we believe theorems? *Philosophia Mathematica*, 17(1), 84–94.
- Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, 7(1), 5–41.
- Rav, Y. (2007). A critique of a formalist-mechanist version of the justification of arguments in mathematicians' proof practices. *Philosophia Mathematica*, 15(3), 291–320.
- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge: MIT Press.
- Robinson, J. A. (2000). Proof = guarantee + explanation. In S. Hölldobler (Ed.), *Intellectics and computational logic* (pp. 277–294). Alphen aan den Rijn: Kluwer.
- Robinson, J. A., & Voronkov, A. (Eds.). (2001). *Handbook of automated reasoning (in 2 volumes)*. Amsterdam: Elsevier.

- Schlimm, D. (2008). Two ways of analogy: Extending the study of analogies to mathematical domains. *Philosophy of Science*, 75(2), 178–200.
- Schlimm, D. (2011). On the creative role of axiomatics. The discovery of lattices by Schröder, Dedekind, Birkhoff, and others. *Synthese*, 183(1), 47–68.
- Schlimm, D. (2013). Axioms in mathematical practice. *Philosophia Mathematica*, 21(1), 37–92.
- Steen, L. A., & Seebach, J. A., Jr. (1970). *Counterexamples in topology*. New York: Holt, Rinehart and Winston.
- Tanswell, F. (2015). A problem with the dependence of informal proofs on formal proofs. *Philosophia Mathematica*, 23(3), 295–310.
- Tanswell, F. (2016). Proof, rigour and informality: A virtue account of mathematical knowledge. Ph.D. thesis, St. Andrews
- Tappenden, J. (2005). Proof style and understanding in mathematics I: Visualization, unification, and axiom choice. In P. Mancosu, K. F. Jorgensen, & S. A. Pedersen (Eds.), *Visualization, explanation and reasoning styles in mathematics* (pp. 147–214). Berlin: Springer.
- Tignol, J. P. (2016). *Galois' theory of algebraic equations* (2nd ed.). Hackensack: World Scientific Publishing, Co. Pte. Ltd.
- Waszek, D. (2018). Rigor and the context-dependence of diagrams: The case of Euler diagrams. In P. Chapman, G. Stapleton, A. Moktefi, S. Perez-Kriz, & F. Bellucci (Eds.), *Diagrammatic representation and inference (Diagrams 2018)* (pp. 382–389). Cham: Springer.
- Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge: Harvard University Press.
- Wussing, H. (1984). The genesis of the abstract group concept: a contribution to the history of the origin of abstract group theory. MIT Press, Cambridge, MA. Translated from the German by Abe Shenitzer and Hardy Grant.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.