

As Thurston Says?

On using quotations from famous mathematicians to make points about philosophy and education.¹

Abstract: it is commonplace in the educational literature on mathematical practice to argue for a general conclusion from isolated quotations from famous mathematicians. In this paper, we supply a critique of this mode of inference. We review empirical results that show the diversity and instability of mathematicians' opinions on mathematical practice. Next, we compare mathematicians' diverse and conflicting testimony on the nature and purpose of proof. We lay especial emphasis on the diverse responses mathematicians give to the challenges that digital technologies present to older conceptions of mathematical practice. We examine the career of one much cited and anthologised paper, WP Thurston's 'On Proof and Progress in Mathematics' (1994). This paper has been multiply anthologised and cited hundreds of times in educational and philosophical argument. We contrast this paper with the views of other, equally distinguished mathematicians whose use of digital technology in mathematics paints a very different picture of mathematical practice.

The interesting question is not whether mathematicians disagree—they are human so of course they do. The question is how homogenous is their mathematical practice. If there are deep differences in practice between mathematicians, then it makes little sense to use isolated quotations as indicators of how mathematics is uniformly or usually done. The paper ends with reflections on the usefulness of quotations from research mathematicians for mathematical education.

Authors:

Prof Gila Hanna, The University of Toronto, 252 Bloor Street West, Toronto, Canada

Dr Brendan Larvor, The University of Hertfordshire, Hatfield, UK

B.p.larvor@herts.ac.uk

+447946433205

¹ We are grateful to Paul Dawkins and three anonymous referees for extensive comments on drafts of this paper.

1 Introduction

The reflections and reminiscences of mathematicians are a rich source of data for mathematics education and the philosophy of mathematical practice. Mathematical research practice is hard to study, because it takes place in private and semi-private places (offices, seminar rooms) and in the minds of mathematicians. In addition, it is technically difficult. For these reasons, we need the reflective testimony of mathematicians to help us to understand what they are doing. On the other hand, the usual reservations about practitioner-testimony apply to mathematics. Adepts in any practice can fail to understand what they are doing, how they are doing it and what conditions make it possible. Moreover, they can disagree among themselves about the meaning and nature of the practice, in which case the testimony of any one practitioner cannot be taken as a reliable guide to the whole.

In this paper, we first review empirical results in cognitive psychology that show the diversity and instability of mathematicians' spontaneous opinions on selected aspects of mathematical practice. Next, we compare mathematicians' diverse and conflicting testimony on the nature and purpose of proof. We lay especial emphasis on the diverse practical responses mathematicians give to the challenges that digital technologies present to older conceptions of mathematics. We examine the career of one much cited and anthologised paper, WP Thurston's 'On Proof and Progress in Mathematics' (1994). This paper has been multiply anthologised and cited hundreds of times in educational and philosophical argument. We contrast this paper with the views of other, equally distinguished mathematicians whose advocacy of digital technology in mathematics paints a very different picture of mathematical practice.

The interesting question is not whether mathematicians disagree—they are human so of course they do. The question is how homogenous is their mathematical practice. If there are deep differences in practice between mathematicians, then it makes little sense to use isolated quotations as indicators of how mathematics is uniformly or usually done.

2 Empirical studies on consensus among mathematicians

Mathematicians certainly disagree about some of the features of mathematics that interest educators and philosophers. For example, mathematicians and philosophers have written about properties of proofs such as elegance and explanatory power (Cellucci, 2015; Lang, 1985; Resnik & Kushner, 1987; Rota, 1997; Steiner, 1978). Many philosophical² papers in this area assume that the proofs that feature in their analyses

² As one referee pointed out, here we see philosophers (rather than mathematicians or educationalists) assuming that practice is homogenous. It does not matter for our argument who it was who provoked the psychologists we cite to test the homogeneity hypothesis empirically. All that matters for our purpose are the methods and the results of the experiments.

As Thurston says...

are obviously elegant (or not) or explanatory (or not). If these properties are obvious features of the proofs, one would expect to see a consensus among mathematicians about which proofs instantiate them. This assumption has recently faced empirical challenges (see especially Inglis & Aberdein, 2016). These empirical studies indicate that consensus among mathematicians regarding the elegance and explanatory power of proofs cannot be assumed.

The qualities of proofs

In a pair of papers, Matthew Inglis and Andrew Aberdein have subjected the assumption of consensus among mathematicians to an empirical test. Their first paper in this area (2015) analysed the adjectives that mathematicians commonly use to describe proofs and found that they form four clusters: aesthetics, intricacy, precision and utility. This work used the same method as the classic studies that identified five dimensions to human personality (Goldberg, 1981). With this result in hand, they were able to test the level of consensus about the nature of a given proof and to check the robustness of judgments by comparing the scores for words in the same cluster. This they did in their second paper (2016).

In this second study, they used a proof of the Sylvester-Gallai theorem taken from Aigner and Ziegler (2000). This proof has some obvious advantages: it is short (so reading it does not take up much of the participants' time), it is non-trivial and, since it is taken from a book of famously neat proofs, there is a good chance that some participants would describe it in decisively positive terms. Their survey asked 112 mathematicians to judge the accuracy of twenty adjectives that might describe the proof, using a five-point Likert scale. Inglis and Aberdein (2016) found that the mathematicians' answers formed three groups. The first group rated the proof high on adjectives associated with aesthetics, precision and utility, but low on intricacy. The second group rated it low on aesthetics, intricacy and precision, but high on utility. The third group rated it low on all dimensions, and especially on aesthetics (p. 171). Statistical tests to check for correlations with area of expertise or career stage were non-significant. It seems that this sample of mathematicians took very different views of the qualities of this particular proof, for reasons that are not revealed by the data. Notably, a majority of them rated it below the midpoint of the aesthetic scale, even though it is taken from a book of proofs selected for their beauty. This is, of course, just one study awaiting replication and it is logically possible that there might have been something atypical about this proof or the method of recruiting participants that

As Thurston says...

produced this result. On the other hand, there has been no outcry among mathematicians about the choice of proofs in Aigner and Ziegler (2000).³

There has been one replication of this study by the same authors (forthcoming). They ran the same experiment, but with an additional feature: half of the participants were told that the proof in the study was taken from Aigner and Ziegler's *Proofs from the book*. In their analysis, Aberdein and Inglis distinguished between pure and applied mathematicians (the latter including statisticians). The information that the proof is in *Proofs from the book* made no difference to applied mathematicians. However, among pure mathematicians, the ones who knew its provenance rated it higher on the aesthetic dimension than the ones who did not (on average). A natural explanation for this is that pure mathematicians are more likely to know that Aigner and Ziegler's book is an attempt to realise Erdős's idea of the Book of perfectly elegant proofs. Knowing that this is a proof that one is supposed to admire for its beauty seems to have nudged them towards admiring it themselves. This should not be a surprise. It means that mathematicians are human and that they modify their opinions to conform with their in-group consensus just like everyone else. On the other hand, it does mean that research into the question of what makes a proof beautiful cannot start by assuming that mathematicians agree which proofs are beautiful, nor that such agreement as there is results simply from mathematicians having insight into the beauty of proofs. Mathematicians are far from unanimous in their judgments about the qualities of proofs, in spite of whatever social pressure they may feel from each other.

One might respond to this result by observing that of the categories in Inglis and Aberdein's list, 'aesthetic' is a matter of taste, or at least of personal judgment and 'utility' is relative to context. Perhaps it is not surprising that such judgments vary among mathematicians just as they do among members of any other professional group. If we were to invite a random sample of architects to judge whether a given structure is elegant or functional, we might expect a similar diversity, not because such judgments are entirely personal but just because there are distinct traditions and schools of thought within architecture. Despite this diversity, there may still be something called 'architectural practice', though we should not expect to learn about it by asking just one architect or by collecting data on the assumption that it is homogenous. What, though, about features of proofs that are not matters of taste? We have already seen that mathematicians can disagree about the precision and intricacy of a proof. What about the safety of inferences? What about rigour? One might expect mathematicians to agree about that, even if they disagree about the

³ One computer scientist did say that he suspected that the number of proofs from graph theory in this book was evidence of a campaign by graph theorists to advance their sub-discipline, but even he did not complain that the proofs selected are too ugly to deserve selection (private communication).

As Thurston says...

beauty, precision, usefulness and intricacy of proofs. We turn now to experiments that test for consensus on proof-validity.

The validity of a proof

In a study (Inglis et al, 2013) with a similar design to the ones on aesthetics, it was found that the mathematicians who responded did not agree when asked whether a given proof was valid, and those who judged it to be invalid gave three different reasons why it was not valid. The proof used in this study is probably not a candidate for the Book, as it is a piece of undergraduate-level real analysis that proceeds by making substitutions and taking limits, for the familiar conclusion that $\int x^{-1} dx = \ln x + c$. Of the 109 participants who completed the survey, 29 (27%) judged the argument valid, and 80 (73%) judged it invalid. The study had some additional features—it analysed the results by area of expertise and gave participants an opportunity to revise their judgment when presented with a reason why the proof might be invalid. For the present purpose, however, all we need is that there was substantial disagreement among expert mathematicians about whether a relatively elementary proof is valid (though, there was a clear $\frac{3}{4}$ majority for judging it to be invalid). It would be too quick to conclude from this that mathematicians disagree about what ‘validity’ means. We do not know how carefully the participants examined the proof, and if they read it quickly, they would be especially vulnerable to the common human tendency to attribute soundness to arguments for conclusions that are already believed to be true. Moreover, a printed proof always has gaps, and is best regarded as a series of instructions for the reader to create a mathematical argument of their own. These instructions can be more or less detailed (the shortest we know of is on page 96 of Awoday (2010) and simply consists of the words ‘diagram chase’). Perhaps the judgment of the 29 mathematicians who thought this proof is valid was that ‘in my hands, this is the recipe for a valid argument’. Perhaps all the mathematicians in this study would agree with a general definition of proof validity even though they disagree about whether this proof satisfies that definition. Nevertheless, even if this study does not give us the exciting conclusion that mathematicians disagree about what ‘validity’ means, it does show that mathematicians cannot be relied on to be unanimous even on apparently technical questions about elementary calculus.

One feature of this study that makes it hard to interpret is that it asked mathematicians to assess an entire proof for validity. We might expect more agreement about specific inferences, which was addressed by the next study.

Graphical inferences

The next group of experiments we consider asks about the validity of specific inferences. According to the dominant view in mathematics and philosophy, graphical or pictorial

As Thurston says...

inferences are not rigorous and strictly speaking have no place in mathematical proofs.⁴ However, they do play an essential role in ancient Euclidean geometry, and Manders (2008a, 2008b) has argued that ancient Euclidean proof-practice is perfectly rigorous. To make this argument, Manders introduced a distinction between metrical and non-metrical information expressed in a diagram⁵. Metrical information is sensitive to small deformations in the diagram (such as measured angles and lengths) whereas non-metrical information is not so sensitive (such as one figure being inside another). Manders' argument is that Euclidean plane geometry is rigorous because its proofs only ever take non-metrical information from its diagrams. This, on his view, is how geometers are able to make good arguments on badly drawn diagrams.

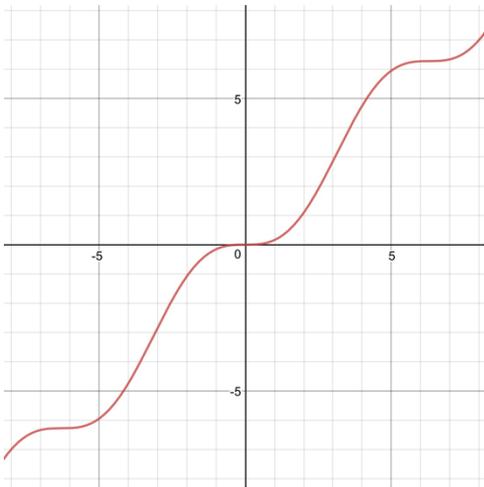
Zhen et al (2016) set out to discover whether students conform to the view that graphical inferences are not rigorous, and their study distinguished between metrical and non-metrical graphical inferences. They found that students do not regard metrical graphical inferences as acceptable, but many of the same students do seem to accept non-metrical graphical inferences. So, it seems that they failed to learn the standard view that graphical inferences of any sort are unrigorous. Moreover, they distinguish between metrical and non-metrical graphical inferences. They must have learned this distinction from somewhere. Weber and Mejia-Ramos (2019) investigated the obvious suspects: mathematics professors. They ran the same experiment, this time with professors as participants rather than students. They got a similar result: very few professors were willing to accept metrical graphical inferences, but a little more than half were willing to accept non-metrical graphical inferences.

⁴ The locus classicus is Bolzano 1817. In the twentieth century, the view that diagrams have no place in proofs is associated with Bourbaki (see Brown 1999 p. 172). Littlewood, presenting himself as resisting the trend, complains that his students “*will not use pictures*” and blames this on “heavy warnings” intended to break students from school mathematics (Littlewood 1953 p.36). Moreover, the rejection of diagrams in proofs is a consequence of the view that the inferences in a mathematical proof should be purely logical, that is, make no reference to specific subject matter. An inference that depends crucially on a diagram obviously violates that rule. This view has its origins in the work of Pasch, Frege and Hilbert. For a recent expression, see Hales 2012 p. x.

⁵ Manders used the terms ‘exact’ and ‘co-exact’. ‘Metrical’ and ‘non-metrical’ are the terms used by Zhen et al, and by Weber and Mejia-Ramos. Since we discuss their argument, we use their terms rather than Manders’.

As Thurston says...

In this experiment, Weber and Mejía-Ramos studied cases from undergraduate real analysis, because here it is common to draw graphs of functions, and one can ask whether these graphs are merely aids to intuition or whether they play a role in proofs. For example, one might look at a graph of $x - \sin x$ and infer that it has a root at zero, that it is strictly increasing, and that it has no other roots (fig. 1). The first two inferences are metrical, in that they rely on the graph being precisely drawn. In order to draw these inferences, we have to know that this is not a slightly deformed graph of



some other function, such as $(1 - \frac{1}{10^{100}})x - \sin(x + \frac{1}{10^{100}})$, but we cannot judge that by inspection of the graph. On the other hand, if we know that the function is strictly increasing, then we can infer from the graph that it has no other roots besides the one at or close to zero. This inference does not depend on the graph being precisely drawn and so is non-metrical.

As always with empirical research, a single study can never be wholly conclusive, and it is easy to think of ways to criticise the experimental set-up.

One point that the authors discuss is that these inferences were presented as happening in a class, so the question is not whether they are rigorous mathematics in some absolute sense but rather whether they are acceptable as undergraduate mathematical reasoning. Since the aim of the study of professors was to compare with the experiment on students, it is reasonable that the same conditions were specified. In some cases, one might wonder whether they are properly described as graphical or pictorial reasoning. Nevertheless, the results pass standard tests for significance, so we can be confident that the experiment has identified something, and that professors are successfully imparting whatever it is to their students. For our purposes, all we need is that there seems to be a disagreement among mathematicians on a question—the nature of rigour—where one might have expected unanimity. One would like to see another experiment of the same sort, except with the proofs presented in a professional research context, so that the question becomes whether either sort of graphical inference is acceptable for publication in a journal or presentation at a conference. Even if the discovered effect were to disappear in that context, we would have discovered a lack of unity in what mathematicians teach their students. We therefore cannot speak univocally of ‘rigour as understood in undergraduate classes’.

3 Mathematicians discuss mathematics

So far, we have only considered the judgments of single, isolated mathematicians (because the experiments so far reported ask participants to give their judgments

As Thurston says...

without discussing the questions with others). But stable mathematical knowledge is the result of a social process. As David Hume put it,

There is no Algebraist nor Mathematician so expert in his science, as to place entire confidence in any truth immediately upon his discovery of it, or regard it as any thing, but a mere probability. Every time he runs over his proofs, his confidence encreases; but still more by the approbation of his friends; and is rais'd to its utmost perfection by the universal assent and applauses of the learned world.⁶

Hume may have underestimated the confidence that some individual mathematicians have in their own judgment. Nevertheless, our interest is in whether mathematical practice is stable and relatively homogenous, or whether it shows the sort of change and variety that is typical of architectural practice. One might reasonably⁷ object to the experiments cited so far that they leave out the stabilising, homogenising effects of interaction between mathematicians. Sure, mathematicians might make different judgments in isolation, but the practice of mathematics includes sharing notes, eliminating each other's errors and critical discussion. We turn, therefore, to public debates about mathematics among mathematicians and in particular to debates about the nature and purpose of mathematical proof. This is because our interest in the testimony of mathematicians concerns the use that philosophers and mathematics educators make of it, so we should look at discussions of the sorts of issues that interest them. Here, then, we encounter the sorts of texts from which philosophers and researchers in mathematics education tend to draw the quotations that they reproduce in isolation and without context.

Testimony of mathematicians

Within the educational and philosophical literature on mathematical proof, one discussion stands out and within that discussion, one paper in particular has become a canonical source of quotations: William Thurston's (1994) 'On proof and progress in mathematics', his response to (Jaffe & Quinn, 1993). This essay has been lifted from its

⁶ Hume (1888), Book I, PART IV. of the sceptical and other systems of philosophy. SECTION I. Of scepticism with regard to reason.

⁷ Matthew Inglis has argued that this objection is not as reasonable as it sounds, because in shifting from cognitive science to ethnography, we are changing the object of study. "Mathematical practice' can be interpreted from an individualist cognitive perspective or an individualist social psychological perspective, or a sociological perspective (roughly speaking I'd say that the difference between the latter two is that the object of study for a social psychologist is the individual in a social setting, whereas the object of study for the sociologist is the social setting itself)... It is like a sociologist saying, 'cognitive experiments don't consider social factors' or a cognitivist saying 'ethnographies don't consider cognitive processing'. Both those statements are true, but they're not sensible objections." (private communication, 2018).

As Thurston says...

original polemical context, reprinted as Thurston (1995) and anthologised in Tymoczko (1998) and Hersh (2006). These four editions of Thurston's article have over 900 Google Scholar citations between them, including in works by the present authors, such as Hanna (1995) and Larvor (2012, 2019). Amongst other things, this article includes a version of the idea we pursue here, i.e. that we should not assume that, "there is uniform, objective and firmly established theory and practice of mathematical proof" (p. 161). Understanding why this one article took off would indicate what philosophers, mathematicians, and mathematics educators find useful in such testimony.

Before considering Thurston's article, it is worth recalling what this debate was about. Jaffe and Quinn's paper of 1993 presented a vision of mathematics in danger. Their article claimed that the distinction between mathematical speculation—including informed, reasoned speculation with plausible arguments—and rigorously proven mathematics has frayed. Their diagnosis had two parts: first, the arrival of ever more powerful computers has allowed the emergence of 'experimental' mathematics, and second, theoretical physics offers a scientific milieu where mathematicians can do serious and interesting mathematical work without having to produce the proofs that mathematics journals would normally demand. Their concern was chiefly not with what is done, but with how it is published. Mathematicians, they insist, should not be allowed to rush into print and claim results that have not been proved in detail, even if they have a sketch of how the proof will go when they get around to writing it out. Aside from the possibility of publishing erroneous claims, they argue, such premature publication reduces the motivation for working out the needed details. If the honour of discovering a result has already been claimed by a senior mathematician who published it along with a proof-strategy or sketch, then there is little incentive for others to fill in the gaps (Jaffe & Quinn, 1993, p. 8). Jaffe and Quinn give various cases that they present as examples of this effect, including work by Thurston. Jaffe and Quinn are not widely quoted in the philosophical and educational literature, for the obvious reason that their image of mathematical proof—in which the formal published document is the paradigm of 'proper proof'—has more distinguished expositors. Anyone looking for a quotation to express or illustrate this view would look elsewhere, to David Hilbert or Bertrand Russell, for example, or to Saunders Mac Lane or Thomas Hales.

The success of Thurston's riposte (at least as measured by citations in philosophical and educational writing) cannot be simply a function of his fame or distinction. After all, fifteen equally famous mathematicians replied to Jaffe and Quinn (Atiyah et al, 1993), but their responses have not been mined for quotations to anything like the same degree. Unlike the other mathematicians under fire from Jaffe and Quinn, Thurston replied with a full-length, wide-ranging essay that started from the question of why anyone should do mathematics at all. His answer focussed on the importance of understanding mathematics, both the proofs and the results. The core of his defence is that the understanding of mathematical ideas is a necessary condition for the formal

As Thurston says...

checking that Jaffe and Quinn care about. One of the points at issue between Thurston and Jaffe and Quinn was that, according to Thurston, the paper record of mathematics published in journals is incapable of sustaining the shared understanding that mathematicians need in order to do their work. Real mathematical understanding depends on real mathematical communication, which for him meant in-person dialogue. In particular, several statements in Thurston's response appealed to researchers in mathematics education who are (justifiably) preoccupied with mathematical understanding (and were so before he published this article—see, for example, Hanna & Winchester, 1990; Hanna, 1989; de Villiers, 1990). Because researchers in mathematics education fear, rightly or wrongly, that excessive emphasis on rigour or on formal proof could hinder comprehension, they find the following statement from Thurston's essay highly quotable:

...we should recognize that the humanly understandable and humanly checkable proofs that we actually do are what is most important to us, and that they are quite different from formal proofs. For the present, formal proofs are out of reach and mostly irrelevant: we have good human processes for checking mathematical validity. (Thurston, 1994, p. 171)

Thurston's essay stands out because it does not merely affirm the importance of mathematical understanding (which no one would deny), but further claims that rigorous proof depends on mathematical understanding. The 'good human processes' he refers to here are only possible because mathematicians have an intimate understanding of the subject-matter of these 'humanly understandable and humanly checkable' proofs. This was the core of his defence of his practice of talking about results in public before he published verified proofs: he had to teach the mathematics community to understand the concepts he was working with. Otherwise, there would be no one able to review and validate his work. A similar situation seems to be holding up a decision on Shinichi Mochizuki's claimed proof of the ABC conjecture. Mochizuki developed a new body of mathematics called Inter-universal Teichmüller theory in order to prove it. So far, the few mathematicians who understand it are mostly Mochizuki's students or associates. This makes it very difficult for the international mathematics community to reach a consensus on the reliability of his proof. Here too, the option of sending the proof for review in advance of presenting it publicly was not available (see Klarreich, 2018).

Thurston's insistence that communal understanding precedes proof-checking makes it attractive to recruit him to the 'understanding first' side of a struggle over mathematics teaching in secondary schools (the other side being the 'procedural first drill' faction)—so long as his remarks are ripped from the context in which he made them. (See Barwell and Abtahi (2017) for the construction of this conceptual vs. rote-learning division in the news media; also see Manin (1998) for further discussion of Thurston's argument.)

As Thurston says...

Thurston's argument is that the mathematical community needs to get itself into a position to review and validate new mathematics by spending time learning it. Peer review assumes the existence of peer reviewers with the appropriate expertise, and when whole new subfields are being created, it takes time and collective work to realise this assumption. This point applies only to research mathematics and then only in the relatively rare circumstance where one individual pioneers a whole new body of thought. It has no obvious direct application or relevance to school classrooms, yet Thurston's paper is quoted by education researchers who are principally interested in school mathematics. It's true that Thurston starts with some general remarks about the importance of understanding in mathematics, but these are a statement of common ground. Everyone agrees that understanding is important. The serious discussion is about its logical and developmental relations with other aspects of mathematical practice: formal proof in research and procedural drill in school. Jaffe and Quinn can retort that it's important to be sure that whatever we try to understand is true. The 'drill first' faction in school education will say that real understanding depends on a firm foundation of familiarity with basic facts and procedures. However, this parallel is misleading. Thurston's argument is about his research community as a whole coming to understand and validate new knowledge, whereas the debate about school mathematics concerns individual students getting to grips with material that was validated long ago. Tying these together logically would require some version of the old (and largely discredited) idea that ontogenesis must recapitulate phylogenesis. As soon as Thurston's argument moves beyond the truism that understanding is desirable, it loses much of its relevance to schoolteaching.

4 The challenges and opportunities of computer mathematics

Examination of the polemical context of Thurston's article has shown that he did not speak for all research mathematicians, knew he did not, and his vision of mathematics cannot be taken as an unproblematic description of a homogenous item called 'mathematical research practice'. In the present section, we consider some of the technological changes that have taken place in the quarter century since Thurston published his article. As we shall see, these have multiplied and deepened divisions among research mathematicians about what constitutes good proving practice. We take no sides in these controversies; our aim here is only to record them.

One of the ambiguities in the Jaffe and Quinn controversy is the sense of 'formal'. Jaffe and Quinn write of 'formal' proof, by which they mean proofs that are sufficiently detailed and regimented for publication in a research journal. However, this is not what logicians or computer scientists mean by 'formal'. Mathematicians writing for mathematics journals do not normally specify their notation in advance or make explicit the strength of the logic of their inferences, as may be done in logic or computer science. Mathematicians consider proofs suitable for publication when they display an

As Thurston says...

organized structure, show the relevance of lemmas and definitions, and provide enough information – sufficient, but not necessarily complete – to convince an informed reader (that is, a trained mathematician) that the conclusion follows from the stated premises.

Reviewers usually apply these same standards when evaluating the merits of a proof submitted for publication. There is even some evidence that the refereeing process is more about deciding which papers are of sufficient interest and novelty than about assessing their correctness – with the result that some published mathematical papers may contain errors (Geist et al, 2010). Over the last decade, some prominent mathematicians have expressed their dissatisfaction with seeing too many reasoning errors in informal proofs (Hales, 2008; Voevodsky, 2014). Whether the proofs are their own or offered by others, they would like to have a higher level of confidence in their correctness than the heretofore acceptable methods can provide, and so are coming to prefer more formal methods—formal, that is, in the sense of formal logic. Computers play a vital role in this development.⁸

A formal proof in this sense, though it may have been inspired by an intuitive grasp of its subject matter, in itself makes no appeal to intuition. Rather, it makes explicit every single step of the derivation, respecting the rules of logic, and thus is less susceptible to subtle logical errors. This is so because a formal proof (in this sense) is a finite sequence of well-formed formulas, each of which is an axiom, an assumption, or follows from the preceding formulas in the sequence by a rule of inference.

Mathematical logic, built upon formal axiomatic deduction, offers just such a model of proof, one that is capable of attaining the desired higher level of rigour, precision, and certainty. But a purely formal proof in mathematical logic, written entirely in a symbolic language, would require an enormous effort to produce, would be much longer, and would be very difficult if not impossible to understand. Thus, logicians face the same difficulty as mathematicians, and like them tend to rely on informal inferences, writing their proofs in a combination of natural and symbolic language that is relatively easy to read and understand.

It was never feasible for practicing mathematicians or logicians to construct a formal proof of any complexity by hand. The advent of computers, however, and recent advances in proof-assistance technology, have made it possible to consider widening the use of formal proof through machine-assisted translation of informal proofs into formal ones. Thurston's claim that "formal proofs are out of reach and mostly irrelevant" (Thurston, 1994, p. 171) is facing a growing challenge. Formal

⁸ One reviewer wondered whether, in this section, we are ourselves guilty of argument by isolated quotation. We don't think so, because we don't take these mathematicians to speak for the whole of mathematics, and besides the argument of this section appeals to a change in mathematical practice. The mathematicians we cite here are doing mathematics differently, and that is what breaks up the homogeneity of current practice.

As Thurston says...

proofs are increasingly feasible and for a growing minority of mathematicians, far from irrelevant.

Mathematicians promoting the use of technology

As mentioned above, the interest of some mathematicians in formal proof stems from a desire for more thorough verification of their own research and from a perception that the conventional refereeing process is far from adequate (Hales, 2008; Harrison, 2008; Geist et al, 2010). As Harrison put it: "We welcome the prospect of formalizing mathematics. In our view, the traditional social process is an anachronism to be swept away by formalization, just as empiricism replaced a similar 'social process'" (Harrison, 2008, p. 1400). Here, then, is a movement of senior mathematicians who are so unimpressed by Thurston's 'good human processes' for proof evaluation that they are going to the considerable trouble of replacing them with machine checking.

With a view to promoting the use of formal methods in mathematics, the *Notices of the American Mathematical Society* published a Special Issue on Formal Proof (Notices, 2008, vol. 55, no. 11). They state that "Using computers in proofs both extends mathematics with new results and creates new mathematical questions about the nature and technique of such proofs. This special issue features a collection of articles by practitioners and theorists of such formal proofs which explore both aspects." This special issue points out that mathematicians now have access to very promising ways to construct formal proofs and thus to gain increased confidence in their correctness.

With the development over the past forty years of computer programs known as "proof checkers" or "proof assistants", checking the correctness of a formalized proof has reached a level that no ordinary proof can match (Avigad & Harrison, 2014). According to Wiedijk (2008), these programs have been successful in checking the validity of the proofs of several well-known theorems, such as the Fundamental Theorem of Algebra (2000), the Jordan Curve Theorem (2005), the Fundamental Theorem of Calculus (1996), the Four Colour Theorem (2004), and the Prime Number Theorem (2008) and more. In addition, computer scientists have added a number of features to proof checkers and designed interactive proof assistants so that computers can now both discover new mathematics and verify the correctness of mathematical claims. These interactive proof assistants also known as "automated theorem provers" can help mathematicians translate their informal proofs into formal ones written in a well-defined logic accepted by the system they are using and can even help mathematicians in the process of reasoning (Bundy, 1991, 2013). Automated proof assistants interact with the mathematician to find a formal proof.

All four contributors to the special issue agreed that the use of formal proof is a trend that will continue and that will contribute to enhancing rigour and gaining confidence in the correctness of long and complex proofs that are not amenable to human

As Thurston says...

verification. They all described in detail how proof assistants help mathematicians attain the utmost level of reliability, adding that in addition to reliability, the formalization of mathematical proofs can be a very worthwhile activity in its own right as it combines computer programming with pure mathematics. Aside from increased rigour, another attraction is that many such programs have generated a large number of new conjectures in several areas of mathematics.

Fields medallist Vladimir Voevodsky was one of the strongest supporters of this third revolution. He found the use of proof assistants to be extremely helpful and even necessary in verifying his own proofs, since they are capable of checking every inference for correctness. Here, he describes his use of computers:

“And I now do my mathematics with a proof assistant ... at least I don’t have to go home and worry about having made a mistake in my work ... nor do I have to worry about my arguments being too complicated or about how to convince others that my arguments are correct. I can just trust the computer. There are many people in computer science who are contributing to our program, but most mathematicians still don’t believe that it is a good idea. And I think that is very wrong.” (Voevodsky, 2014, p. 9.)

Ten years after the publication of the special issue on formal proof in 2008, the *Notices of the AMS* published a paper by Jeremy Avigad, “The Mechanization of Mathematics” (2018) that could be considered as a follow-up to this issue. Avigad recounts the development of formal methods in computer science and their application to mathematics, describing how they can be used both for mathematical verification and for mathematical discovery. Referring to the current state of mathematical practice, he concedes that “it may seem premature to predict that formally verified proof will become common practice” p. 685.

Regarding the role of the new technologies in mathematical discovery, Avigad notes that the use of formal methods in discovery is even less advanced than their use in verification. He attributes this lag to the fact that mathematicians may not yet have enough experience with using formal methods in mathematical discovery to be able to appreciate their power. He hopes that in the course of time, mathematicians will be

... open to the possibility that new technologies can open new mathematical vistas and afford new types of mathematical understanding. The prospect of ceding a substantial role in mathematical reasoning to the computer may be disconcerting, but it should also be exhilarating, and we should look forward to seeing where the technology takes us.” (Avigad, 2018, p. 688).

Bundy (2011) cites a few reasons given by mathematicians for not using automated provers, among them the following: automated provers are too hard to use, they are

As Thurston says...

not powerful enough to prove novel conjectures, and why give up the fun of proving? (p. 13). According to Ganesalingam and Gowers (2017), the main obstacle to the wider use of automated assistants seems to be that currently available systems cannot do mathematics in a “human style”. They believe that a theorem prover capable of producing proofs that ‘explain what is going on’ (unfortunately, not in existence yet) would likely be more appealing to mathematicians. Nevertheless, a sub-community of mathematicians does use proof assistants (also known as interactive theorem provers). Perhaps the majority of mathematicians are waiting for automated assistants to be more user-friendly and to build a track record in generating conjectures -- and perhaps even to take a qualitative step from assistance with formal proofs to the validation and construction of heuristic arguments. For our purposes, the significance of these quoted opinions is that they are trying to understand the pace of change in the practice of mathematical research (and why it is not quicker). That some very distinguished mathematicians have changed the way they work is not in dispute.

These observations return us to Thurston’s claim, as quoted above:

...we should recognize that the humanly understandable and humanly checkable proofs that we actually do are what is most important to us, and that they are quite different from formal proofs. For the present, formal proofs are out of reach and mostly irrelevant: we have good human processes for checking mathematical validity.” (Thurston, 1994, p. 172)).

The last sentence is no longer true, if it ever was. We do not seem to have good human processes for checking the validity of Mochizuki’s purported proof of the ABC conjecture (to name but one example), and formal proofs of some theorems (‘formal’ in the sense of logic and computer science) are within reach and some mathematicians find them relevant. Mathematical practice changes, it is still changing, and (in spite of the impression that one may gain from some history books) it does not change uniformly, with all mathematicians marching in step. New challenges and opportunities, in this case from technology, introduce further heterogeneity.

5 Concluding thoughts

Citing a text is a paradigmatic move in humanities disciplines, and should be carried out with appropriate rigour. This means paying attention to the context of the quoted lines. To take our leading example, Thurston’s paper is often quoted as if it were a collection of disinterested observations by a senior mathematician who has decided to share the collective wisdom of top researchers in his field. In fact, in this paper Thurston is defending his professional practice against a very aggressive attack. Jaffe and Quinn effectively accused him of claiming honours that he did not deserve, of having presented as proved theorems what were, then, no more than conjectures supported by vague proof-sketches. There are few more serious crimes in mathematics. Notice

As Thurston says...

that philosophers and educational researchers who quote Thurston do not usually quote the part where he advises us not to assume that, “there is uniform, objective and firmly established theory and practice of mathematical proof” (p. 161). This, after all, is precisely the assumption that one must make in order to lift a quotation from Thurston’s article and present it uncritically as part of the truth about mathematics. Knowing the polemical context of Thurston’s article reminds us that it is not a description of mathematical practice, but rather an argument about mathematical practice. This puts one on the alert to look out for its premises, in case they do not carry over into the context where a quotation from it might be deployed. We have already seen this: strictly speaking, by its own logic, his argument applies only to pioneering mathematical research that runs ahead of the capacity of the mathematical community to assess proofs that use its new ideas. Thurston does not state this restriction, but he was careful to specify that his argument may only apply to his area of mathematics, because other fields may have less use for spatial and kinematic intuition. Rigour, in the humanities disciplines that rely on textual quotation, means taking proper account of the context from which the quotation is taken. Doing this with Thurston’s article dramatically reduces its relevance to mathematics education.

There are, of course, innocent ways of using a quotation. One might quote an author simply because they have captured a point with unusual elegance, pith or wit. One might quote a text because it is the standard reference point for the thought one wishes to discuss (and this may be one reason why Thurston is so widely quoted—he has become canonical). One might quote a text in order to refute it (in philosophy, one of the most common reasons to quote an author is to show that they really did commit to the thesis that one is about to attack). A quoted opinion from a senior mathematician can supply a hypothesis for researchers, who can set out to find out how widely the opinion is shared. Quotations have an obvious place in biographies and in history—if used with the standards of rigour that apply to such writing. Quotations from senior mathematicians can suggest interpretations of raw empirical data on mathematical practice, so long as the interpretation does not rely on the seniority of the mathematician for its plausibility. Finally, one might quote a text as expert testimony, but this is the weakest and slipperiest reason, because in philosophy and education there is almost certainly another expert of similar standing who holds the opposite view. In these areas of enquiry, having an expert commit to an opinion in print does not establish that the opinion is true. It only establishes that it is discussable. Without embarking on a survey of the 900+ citations of Thurston’s article, we cannot say for sure how many writers on mathematics education are guilty of using Thurston as an authority on the nature of mathematics in general. Nevertheless, the fact that his article is quoted orders of magnitude more frequently than everyone else does suggest that much of this mass of quotations may be more rhetorical than rigorous (because

As Thurston says...

rigorous examination of expert opinion would require quotations on a similar scale of voices disagreeing with him).

Beyond this narrow point about textual quotation, there is a larger question about the unity of mathematical practice. The interesting question is not whether mathematicians disagree—they are human so of course they do. The question is: how homogenous is their mathematical practice? If there are deep differences in practice between mathematicians, then it makes little sense to use isolated quotations as indicators of how mathematics is uniformly or usually done. We saw that studies in cognitive science suggest that mathematicians disagree about the aesthetic and logical features of proofs. Furthermore, we have seen that mathematicians do in fact disagree explicitly about the nature and value of proofs and the validation process for mathematical research. One might, therefore, be tempted simply to say that there is no such thing as mathematical practice. However, this would be premature. First, our evidence for diversity in mathematical norms was the result of extensive searches, and was found in relatively obscure corners such as mathematicians' views about graphical inferences or the very frontiers of technological developments where the discipline is still deciding what it thinks. It would be an error to focus on these cases and ignore the great swathes of normative judgment where mathematicians agree among themselves. The analogy with architecture is instructive here: there are schools and trends in architectural practice that prevent one from announcing on a slender evidential basis that architects think this or that. Nevertheless, buildings and theorems have to meet more exacting standards than simply being fashionable or well-regarded.

References

- Aigner, M. & Ziegler, G. (2000). *Proofs from the book*. (2nd ed.) Berlin: Springer.
- Atiyah, M. et al. (1994). 'Responses to "Theoretical Mathematics": Towards a cultural synthesis of mathematics and theoretical physics, by A. Jaffe and F. Quinn', *Bulletin of the American Mathematical Society*, **30**(2), 178–207.
- Avigad, J. (2018). The mechanization of mathematics. *Notices of the [American Mathematical Society](#)*, **65**(06), 681-690.
- Avigad, J., & Harrison, J. (2014). Formally verified mathematics. *Communications of the ACM* **57**(4), 66-75.
- Awoday, S. (2010). *Category theory* (2nd ed.). Oxford: Oxford University Press
- Barwell R., & Abtahi, Y. (2017). Mathematics concepts in the news. In de Freitas, E., Sinclair, N., & Coles, A. (Eds.). *What is a mathematical Concept?* (pp.175-188). Cambridge: Cambridge University Press. doi:10.1017/9781316471128
- Bolzano, B. (1817). Purely analytic proof of the theorem that between any two values which give results of opposite sign there lies at least one real root of the equation, in Ewald, W. (Ed.), (1996). *From Kant to Hilbert. A source book in the foundations of mathematics*, vol. 1, (pp. 225–248). Oxford: Clarendon Press.
- Brown, J. R. (1999 & 2005). *Philosophy of mathematics: The world of proofs and pictures*. New York: Routledge.
- Bundy, A. (1991). A science of reasoning. In J. L. Lassez & G. Plotkin (Eds.), *Computational logic: Essays in honor of Alan Robinson* (pp. 178-198). Cambridge, MA: MIT Press.
- Bundy, A. (2011). Automated theorem provers: A practical tool for the working mathematician? *Annals of Mathematics and Artificial Intelligence*, **61**(1), 3–14.
- Bundy, A. (2013). The interaction of representation and reasoning. *Proceedings of the Royal Society A Mathematical, Physical and Engineering Sciences*, **469**(2157), 1–18. The Royal Society.
- Cellucci, C. (2015). Mathematical beauty, understanding, and discovery. *Foundations of Science*, **20**(4), 339–355, doi: [10.1007/s10699-014-9378-7](https://doi.org/10.1007/s10699-014-9378-7)
- De Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras* **24**, 17-24.
- Ewald, W. (Ed.). (1996). *From Kant to Hilbert. A source book in the foundations of mathematics*, Volumes 1 and 2. Oxford: Clarendon Press.
- [Ganesalingam, M., & Gowers, W. T. \(2017\)](#). A fully automatic theorem prover with human-style output. *Journal of Automated Reasoning*, **58**, 253–291.
- Geist, C., Löwe, B., & Van Kerkhove, B. (2010). Peer review and testimony in mathematics. In B. Löwe & T. Müller (Eds.), *Philosophy of mathematics: Sociological aspects and mathematical practice* (pp. 155-178). London, UK: College Publications.
- Goldberg, L.R. (1981). Language and individual differences: the search for universals in personality lexicons. In Wheeler (Ed.), *Review of Personality and social psychology* (pp. 141-165). Beverly Hills, CA: Sage.
- Hales, T. (2008). Formal proof. *Notices of the [American Mathematical Society](#)*, **55**(11), 1370–1380.
- Hales, T. (2012). *Dense sphere packings: A blueprint for formal proof*. Cambridge: Cambridge University Press.
- Hanna, G. (1989). More than formal proof. *For the Learning of Mathematics*, **9**(1), 20-25.

- Hanna, G. (1995). Challenges to the Importance of Proof. *For the Learning of Mathematics*, 15(3), 42-50.
- Hanna, G., & Winchester, I. (Eds.). (1990). *Creativity, thought and mathematical proof*. Toronto: Ontario Institute for Studies in Education.
- Harrison, J. (2008). Formal proof—theory and practice. *Notices of the [American Mathematical Society](#)*, 55(11), 1395–1406.
- Hersh, R. (Ed.) (2006). *18 unconventional essays on the nature of mathematics*. New York: Springer.
- Hume, D. (1888). Selby-Bigge, L.A. (Ed.). *A treatise of human nature*. Oxford: Clarendon Press.
- Inglis M, Mejia-Ramos JP, Weber K, & Alcock L. (2013). On mathematicians' different standards when evaluating elementary proofs. *Topics in Cognitive Science*, 5(2), 270-82.
- Inglis, M., & Mejia-Ramos, J.P. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction*, 27(1), 25-50.
- Inglis, M., & Aberdein, A. (2015). Beauty is not simplicity: an analysis of mathematicians' proof appraisals. *Philosophia Mathematica*, 23(1), 87 – 109.
- Inglis, M., & Aberdein, A. (2016). Diversity in proof appraisal. In B. Larvor (Ed.), *Mathematical cultures: The London meetings 2012-2014* (pp. 163-179). Basel: Birkhäuser Science.
- Inglis, M., & Aberdein, A. (Eds.). (2019). *Advances in experimental philosophy of logic and mathematics*. London: Bloomsbury Academic.
- Inglis, M., & Aberdein, A. (forthcoming) 'Testing hypotheses about mathematical practice: Are aesthetic judgments in mathematics purely aesthetic?'
- Jaffe, A. & Quinn, F. (1993). "Theoretical mathematics": Toward a cultural synthesis of mathematics and theoretical physics. *Bulletin of the American Mathematical Society* 29(1) 1-13.
- [Klarreich, E.](#) (2018). [Titans of mathematics clash over epic proof of ABC conjecture](https://www.quantamagazine.org/titans-of-mathematics-clash-over-epic-proof-of-abc-conjecture-20180920/#). *Quanta Magazine*. <https://www.quantamagazine.org/titans-of-mathematics-clash-over-epic-proof-of-abc-conjecture-20180920/#>
- [Lang, S.](#) (1985). *The beauty of doing mathematics: Three public dialogues*. New York: Springer-Verlag.
- Larvor, B. (2012). How to think about informal proofs. *Synthese* 187(2), 715-730.
- Larvor, B. (2019). From Euclidean geometry to knots and nets. *Synthese* 196(7), 2715–2736.
- Littlewood, J.E. (1956). *A Mathematician's miscellany*. London: Methuen.
- Manders, K. (2008a). Diagram-based geometric practice. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 65–79). Oxford: Oxford University Press.
- Manders, K. (2008b). The Euclidean diagram. In P. Mancosu (Ed.), *The philosophy of mathematical practice* (pp. 80–133). Oxford: Oxford University Press.
- Manin, Yu. (1998). Truth, rigour, and common sense. In H. G. Dales & G. Oliveri (Eds.), *Truth in mathematics* (pp. 147-159). Oxford: Oxford University Press.
- Nathanson, M. B. (2008). Desperately seeking mathematical truth. *Notices of the American Mathematical Society*, 55(7), 773.
- Pollatsek, H. (2018). How mathematics research journals select articles. *Notices of the [American Mathematical Society](#)*, 65(1), 63-64.
- Resnik, M.D., & Kushner, D. (1987). Explanation, independence and realism in mathematics. *British Journal for the Philosophy of Science*, 38, 141-158.

As Thurston says...

- Rota, G-C. (1997). The phenomenology of mathematical beauty. *Synthese* 111(2), 171-172.
- Steiner, M. (1978). Mathematical explanation. *Philosophical Studies*, 34, 135-151.
- Thurston, W. (1995). On proof and progress in mathematics. *For the Learning of Mathematics*, 15(1), 29-37.
- Thurston, W. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society* 30(2), 161–177.
- Tymoczko, T. (Ed.). (1998). *New directions in the philosophy of mathematics: An anthology*. Princeton, New Jersey: Princeton University Press.
- Voevodsky, V. (2014). The origins and motivations of univalent foundations. *The Institute Letter*, 8-9. <https://www.ias.edu/sites/default/files/documents/publications/ILsummer14.pdf>
- Weber, K., & Mejía-Ramos, J.P. (2019). An empirical study on the admissibility of graphical inferences in mathematical proofs. In M. Inglis & A. Aberdein (Eds.), *Advances in experimental philosophy of logic and mathematics* (pp. 123-144). London: Bloomsbury Academic.
- [Wiedijk, F. \(2008\). Formal proof –getting started. *Notices of the American Mathematical Society*, 55\(11\), 1408-1414.](#)
- Zhen, B., Weber, K., & Mejia-Ramos, J. P. (2016). Mathematics majors' perceptions of the admissibility of graphical inferences in proofs. *International Journal of Research in Undergraduate Mathematics Education*, 2(1), 1–29.