What Is Mathematical Rigor?

John P. Burgess and Silvia De Toffoli

Rigorous proof is supposed to guarantee that the premises invoked imply the conclusion reached, and the problem of rigor may be described as that of bringing together the perspectives of formal logic and mathematical practice on how this is to be achieved. This problem has recently raised a lot of discussion among philosophers of mathematics. We survey some possible solutions and argue that failure to understand its terms properly has led to misunderstandings in the literature.

TABLE OF CONTENTS
1. INTRODUCTION
2. THE ORIGIN OF THE PROBLEM
3. THE PROBLEM AS IT NOW CONFRONTS US
4. SUB-PROBLEMS
5. CONCLUSION
1. Introduction

Mathematics is distinguished from other disciplines by its method. Where empirical science relies on experiments, mathematics relies on proofs. Thus Hempel:

The nature of mathematical truth can be understood through an analysis of the method by means of which it is established. On this point I can be very brief: it is the method of mathematical demonstration, which consists in the logical deduction of the proposition to be proved from other propositions, previously established. Clearly, this procedure would involve an infinite regress unless some propositions were accepted without proof; such propositions are indeed found in every mathematical discipline which is rigorously developed; they are the axioms or postulates (we shall use these terms interchangeably) of the theory. (Hempel 1945, 7)

But can one really be that brief? We will suggest not. If we give the name “rigor” to the quality that distinguishes genuine demonstrations or proofs from flawed purported proofs (and avowedly merely heuristic plausibility arguments), then the characterization of what rigor, and hence genuine proof, consists in is a substantial philosophical problem with multiple ramifications.

Rigorous proof is supposed to guarantee that the premises invoked imply the conclusion reached, and the problem of rigor may be described as that of bringing together the perspectives of formal logic and mathematical practice on how this is to be achieved. To begin with logic, it articulates a semantic conception of what it is for premises to imply a conclusion: they do so if, regardless of originally intended interpretations, any reinterpretation that makes the premises true makes the conclusion true. For instance, reinterpreting “point of the plane” as “point of the unit disc” and “straight line” as “arc orthogonal to the unit circle” as in the Poincaré model of hyperbolic geometry, shows that the other postulates of Euclidean geometry do not imply the parallel postulate.

This conception of what rigor is supposed to guarantee is generally accepted, as is the claim that present-day mathematicians’ proofs (unlike many past mathematicians’ “proofs”) do guarantee implication of the conclusion by the premises. Hence assuming the truth of generally accepted axioms, they guarantee the truth of the theorems, and the correctness or
reliability of mathematics. One consequence is that if logicians show that some proposition is not implied by accepted axioms, mathematicians may take warning that they will be wasting their time if they attempt to give a proof of the proposition in question.

Logic also articulates a syntactic conception of a derivation of a conclusion from premises, or rather, several demonstrably equivalent notions of derivation in different formats, along with a soundness theorem, to the effect that formal derivability guarantees implication, and a completeness theorem, to the effect that if implication holds, a formal derivation exists. It follows that if a genuine mathematical proof is given (thus guaranteeing implication) a formal derivation exists, or could in principle be given.

However, what are published in mathematics journals as rigorous proofs are very far from being the kind of formal derivations that are the logicians’ gold standard, with every step explicit and mechanically checkable. They are not cast in any of the logicians’ formalisms, and involve all sorts of shortcuts. They make use of different types of high-level inferences appealing to background knowledge shared by their intended audience, in the process skipping over what would be many, many steps in any kind of formal derivation. They may make use of diagrams in ways that cannot be in any very direct way translated into sequences of formulas such as appear in formal derivations. They may in places offer only stage directions, for instance, telling us “case 2 is similar to case 1” and then leaving that case to the reader; or they may invoke the expression without loss of generality (WLOG) to indicate that although they have proved a result only in a particular case, their proof holds mutatis mutandis for all cases. The puzzle is thus: how do mathematicians' informal proofs manage to guarantee the existence of a logicians’ derivation or formal proof without being one? How do mathematicians’ and logicians’ notions of provability manage to coincide, when their notions of proof are so different?

There is a partial candidate answer sometimes called the standard view: A mathematical argument is a rigorous proof if and only if it can in principle be converted into a formal derivation.

1 WLOG arguments often implicitly invoke some kind of simple meta-theoretic argument from symmetry, perhaps to the effect that any proof that works in the case $a > b$ will also work in the case $b > a$ simply by rewriting it switching the two letters.

2 (Hamami 2019).
The weakness of the view is that formalizability, or convertibility into a formal derivation, is a vague and controversial notion, leaving much unspecified in several directions. What does it even mean to say that mathematicians’ proofs can be in principle converted into logicians’ derivations? Are there specific agents that should be able to perform such conversion?

In the last few decades, the philosophy of mathematical practice, a class of approaches in philosophy of mathematics focusing on the reality of what mathematicians do rather than idealizations thereof, has grown considerably. The gap between formal derivations and informal proofs is one of the aspects of mathematical practice that philosophers associated with this tendency have taken up and begun to analyze in different ways, with some even challenging the idea that rigor should be understood in terms of formalizability at all. Specific subtopics addressed have ranged from the role of computers to that of diagrams and beyond.

In this article, we aim at mapping the problem of rigor. After offering a brief reconstruction of the history of the problem (Section 2), we discuss how the problem confronts us today (Section 3). We then list several subproblems (Section 4) and sum-up our considerations (Section 5).

2. The Origin of the Problem

In (very) general terms, it is plausible to say that modern mathematics inherited from earlier times two bodies of material, roughly geometric and algebraic, which can be represented by Euclid's Elements and al-Khwarizmi's al-Jabr wal-Muqabalah. The former but not the latter was organized in the form of a deductive science.

The ideal of such a science is that conjectures, however discovered, are not ranked as theorems until justified by rigorous proofs showing them to be implied by previous results and ultimately postulates agreed in advance. But what implication amounts to is the central question of logic, and the ideal could not be more precisely articulated so long as the only logic available was traditional syllogistic.

In any case, the ideal was imperfectly realized even in geometry, and hardly attempted on the algebraic side of mathematics, where early modern work, from the solution to cubic equations to the development of calculus, was characterized more by daring innovations (such as imaginaries and infinitesimals).
Standards of rigor were raised in branch after branch of mathematics at an increasing pace through the nineteenth and early twentieth century, and a comprehensive rigorous reconstruction of the whole of modern mathematics, starting from the Zermelo-Fraenkel axioms for set theory, was undertaken by the Bourbaki group in the 1930s.

Meanwhile, the semantic conception of logical implication had emerged from work on the parallel postulate among other things, and developed form in Hilbert's (1899) Foundations of Geometry, though there were refinements by Tarski in the 1930s. Syntactic notions of deducibility (demonstrably equivalent ones) appeared in Frege’s (1879) Begriffsschrift and Russell & Whitehead’s (1910) Principia Mathematica, volume 1. Soundness was proved, and completeness conjectured by Hilbert in the 1920s, and completeness proved by Gödel in 1929. That a formal deduction from the axioms of set theory exists for any rigorously-proven contemporary mathematical result had become a tenable thesis by the 1940s or so. For instance, the work of Gödel in the 1930s and Cohen in the 1960s, showing that there is no formal derivation either of the continuum hypothesis or its negation from ZFC was generally, uncontroversially accepted as showing that there is no proof or disproof by accepted mathematical standards.

However, despite the conclusion that in principle formal derivations exist for all rigorously-proved mathematical results, no one exhibited any such derivation for any non-trivial result prior to the development of automated proof-assistants in the late twentieth century. Notoriously, it took Russell and Whitehead 379 pages to give a formal proof of $1+1=2$ in the first volume of Principia Mathematica (1910).

From this history it should be apparent why the problem of rigor as we confront it today could not have been recognized much prior to the middle of the last century. Even then the issue of characterizing rigor did not draw much attention. Mathematicians were satisfied that they knew it when they saw it, with only occasional flurries of disagreements – the theoretical mathematics debate of the mid-1990s, and doubts attending the first, pioneering computer-assisted proofs. Philosophers generally had other

---

3 However, once we recognize the problem, we can also look backward and use the tools of contemporary logic to analyze ancient practices (Graziani 2020). For a concrete example, consider the formal system developed for Euclid’s Elements (Avigad, Dean, and Mumma 2009).

4 The theoretical mathematics debate started with a paper by Jaffe and Quinn (1993). It then developed through a series of replies among which the influential (Thurston 1994).

5 See, for example, (Tymoczko 1979). For an explanation by Apple and Haken of why their proof is in good standing, see (Appel and Haken 1986).
concerns, less about the internal workings of mathematics than about its place among other sciences.

Many were especially preoccupied with the *ontological* issue of what room there could be for the abstract objects of mathematics – about whose intrinsic nature (as opposed to their interrelationships with each other) mathematics tells us so little – in a naturalistic picture of the world that seemed otherwise to accommodate only concrete physical particles and forces.

3. The Problem as It Now Confronts Us

The *philosophy of mathematical practice* of the later decades of the twentieth century represents a turn to looking at the inner workings of mathematics. It began with the work of a few isolated precursors, but gained momentum and now comprises a large, heterogeneous body of works (Mancosu 2008; Carter 2019).

One of its characteristics is openness to interdisciplinary collaborations, bringing philosophy of mathematics into dialogue with such other disciplines as history of mathematics, cognitive science, and sociology, just as philosophy of mathematics in the late nineteenth and early twentieth century had been intimately connected with the emerging disciplines of modern logic and computer science.

It is not surprising, then, that the problem of rigor soon came into focus, and that the achieved results of mathematics and logic were found to leave important philosophical questions still open, most especially the question of how best to describe the relationship between informal proofs and formal derivations: Can all informal proofs be *formalized*? More crucially, can their correctness be explained in terms of such formalization(s)?

Being interested in how mathematics is practiced, rather than in an idealization thereof, philosophers of mathematical practice tend to concentrate their efforts on the study of informal proofs and to emphasize the distance between them and formal proofs. This *distance* can be measured in various ways. For example, it can be evaluated in terms of the cognitive abilities that are at play when going through a proof (Giaquinto 2007) or in terms of the number of steps needed to prove a particular result (Pelc 2008).

---

6 See, for example, (Lakatos 1963; Aspray and Kitcher 1988).
Although emphasizing the distance between informal and formal proofs does not commit one to any particular position concerning how the correctness of informal proofs is best understood, some scholars have denied that it should be cashed out in terms of formal proofs. As a matter of fact, there is a divide in the current literature between the proponents of the standard view, and the opponents, who strive to produce an account of rigor without appealing to formal proofs.\(^7\)

We believe that this divide, this \textit{formalizability debate}, is largely misguided, because the very meaning of \textit{formalizability} is ambiguous. On scrutiny, apparently opposing views are not incompatible.\(^8\) The confusion stems from the fact that standard view is itself not a single view, but rather a family of views. Recalling our earlier formulation of the standard view (equating rigor with formalizability), it should be noted that if this thesis holds, we can in general associate to a single proof not just one specific but rather \textit{multiple} formal proofs – logicians having developed many very different-looking but demonstrably equivalent formats for derivations.\(^9\)

In picking out a particular position among the many versions of the standard view, we encounter several choice points. Specific positions are committed to stricter or looser connections between proofs and formal proofs, across several (not necessarily independent) dimensions. For reasons of space, we have to limit ourselves to succinct mention of the choice points that seem most relevant to us. Moreover, for ease of discussion, we will present just two alternatives at each choice point, though there might be more nuanced answers.

\textbf{A) How are the steps of the informal proof and the ones of its formal counterpart related?}

\textit{(A.1)} The formalized versions of the steps of the informal proof are components of the formal proof. Therefore, the process of formalization is (apart from transcription into special symbols) one of filling in.\(^{10}\)

\(^7\) This debate started with a series of papers by Rav and Azzouni (Rav 1999, 2007; Azzouni 2004, 2017). Versions of the standard view are argued for in (Burgess 2015; Avigad 2021; Hamami 2019; Tatton-Brown 2021) while (Tanswell 2015; Leitgeb 2009; De Toffoli and Giardino 2016; Larvor 2012) level critiques to different aspects of the standard view.

\(^8\) Avigad (2021) also tries to reconcile the two camps.

\(^9\) That we can associate different formal proofs to the same informal proof is, in general, not a problem. Tanswell (2015) has, however, used this point to challenge Azzouni’s specific version of the standard view.

\(^{10}\) See (Steiner 1975).
(A.2) Parts of the informal proof, such as passages of reasoning with diagrams, though convertible into a sequence of formal steps, are not themselves expressible as informal counterparts of propositions of any of the usual formal systems.\(^{11}\)

A version of (A.1) was tentatively endorsed by one of the authors, who explicitly excluded the analysis of diagrams from his account of rigor, admitting, however, that it is an important and difficult one (Burgess 2015). The other author’s analysis of rigor starts off precisely by tackling such a thorny issue. In (De Toffoli 2021) it is argued that the presence of diagrams in contemporary mathematics puts pressure on (A.1). Since any satisfactory account of rigor cannot exclude a whole class of proofs that are deemed rigorous by mathematicians, this is evidence for (A.2).

This choice point has to do with comparing specific steps of different proofs. One related problem is the one of determining when are two proofs different and when they are the same. This is a vexed issue for both informal and formal proofs. In (De Toffoli forthcoming) it is argued that, given plausible criteria of identity for proofs, diagrams can be essential to certain proofs – thus providing additional support to (A.2).\(^{12}\)

**B) Is the formalization actual or merely potential?**

(B.1) The process of formalization starts from the axioms. Formalizability is conceived as the process of creating a specific formal proof from given starting points.

(B.2) The process of formalization starts from accepted results. Formalizability is thus conceived as merely potential formalizability in one system or another.

If rigor calls for formalization of an actual deduction from the axioms, presumably those of set theory, then we must first define the natural numbers set-theoretically. But how? Zermelo tells us \(2 = \{\emptyset\}\), von Neumann tells us \(2 = \{\emptyset, \{\emptyset\}\}\), yet there is nothing in mathematical practice after the basic laws of arithmetic have been established that favors one over the other. This is the problem of *multiple reductions*, made famous by Benacerraf (1965), which applies to mathematical objects of all kinds, and

\(^{11}\) See (De Toffoli 2021).

\(^{12}\) See also (De Toffoli 2022).
which occupies a large place in ontologically-oriented philosophy of mathematics.

But if rigor only calls for formalization of a deduction from accepted results, including the basic laws of arithmetic, we need never consider which route may have been taken to get from the axioms to those previously established results. As one of the authors has argued (Burgess 2015), the appearance of some ontological mystery about the number 2 is the result of going the wrong way at this crucial choice point, assuming (B1) rather than (B2).

\[ C) \text{ Is the conversion a “routine translation”?} \]

(C.1) Converting a proof into an informal proof is done by a routine translation.\(^\text{13}\)

(C.2) Converting a proof into an informal proof requires more than a routine translation.

The expression “routine translation” has been interpreted in different ways. Non-technically, to say that a translation is routine means to say that is it simple or mechanical, not involving creative steps. According to Mac Lane, a proof can be routinely translated into a formal proof only if there is an algorithmic procedure to decompose all its inferential steps into basic logical steps (Hamami 2019).

Some opponents of the standard view argued for (C.2). As it is clear from recent efforts on the formalization of mathematics, converting a proof into a formal proof is a major enterprise that often requires inventiveness (Hales 2008).

This choice point relates to the question of who actually performs the conversion. Since in order to go through informal proofs, a vast amount of background knowledge is needed, the conversion is not something that can be performed mechanically. As Steiner (1975) suggested, we should imagine a team composed of mathematicians and logicians. However, since in recent years formalizations are done with specific software (interactive proof assistants – more below), we can add computer scientists to the team.

Note that the answers to these three questions might not be independent. For example, one could argue that (A.2) and (C.1) are incompatible. Some of the works that have been considered to oppose the

\(^{13}\) This is Mac Lane’s (1986) view.
standard view have generally only opposed a specific position in the family and therefore might be reconciled with other specific positions.

The problem of rigor is particularly relevant today because contemporary mathematics is ever more sophisticated, and it is pushing the boundaries of what proofs are. For example, the existence of computer-assisted proofs is something that cannot be ignored. Moreover, some accepted proofs are so big that require large-scale collaborations. More mundanely, some fields are so technical that checking proofs has become a very difficult task. And consequently, some proofs are rarely checked: “A technical argument by a trusted author, which is hard to check and looks similar to arguments known to be correct, is hardly ever checked in detail,” said Fields medalist Voevodsky. It is exactly this lack of quality-control that is responsible for some of Voevodsky’s own incorrect arguments having been for a time widely accepted. This motivated him as well as other mathematicians and computer scientists to develop technological tools to verify proofs (Voevodsky 2014). These are interactive proof assistants, software that can be used by mathematicians to create formal counterparts of their proofs (Avigad 2018).

The standards of rigor became more stringent in time. Most pre-20th century mathematics would not meet the current bar on rigor. But is the standard now set once and for all? Perhaps in the future, the bar will rise again. It is possible that in some areas of mathematics, formal proofs would be requested alongside traditional ones.

4. Sub-Problems

In this section, we briefly discuss some sub-problems of the general problem of mathematical rigor. To be sure, this is not an exhaustive list, and there is more to say about each item, but it is at least a start.

4.1. What Is the Function of Rigor?

One function of rigor is to guarantee the correctness of results: if a proof establishes that earlier results and ultimately accepted axioms do genuinely imply a new result, then presuming the former are true, the latter will be so

---

14 The most famous example of such proofs is given by the classification of finite simple groups theorem. For a discussion of this case from the point of view of social epistemology, see (Habgood-Coote and Tanswell 2021).
as well. And indeed a large part of the impetus for historical phenomenon
rigorization came from doubts about the reliability of older results and
methods. However, this cannot be the whole story because there are some
non-deductive arguments in mathematics that do not pretend to be rigorous
but are very reliable. If ensuring correctness is the only function, then there
are different types of rigor (see next question). The function of rigor is
perhaps to guarantee not only correctness, but that there is a specific path
from the starting point to the result. Another, possible function of rigor is to
resolve disagreements. And in fact we observe that disagreement in
mathematics is only rarely recalcitrant.  

4.2. Are there different types of rigor?

Francesco Severi, one of the leading exponents of the classical Italian
school of algebraic geometry, distinguished between substantial and formal
rigor. The former consists in faithfulness to the mathematical facts, while
the latter is understood in connection with formalization. In his words,

rigor is completely satisfied when we say, as in the oath to the
tribunals, the whole truth and nothing other than the truth: a rigor
that I could call substantial to distinguish it from the formal rigor
which calls each time an axiomatic systematization from the letter A
to the letter Ω. (Severi 1949, 2)

4.3. What is, if any, the difference between rigor and acceptability?

As an alternative to the standard view, some have tried to characterize rigor
as what is accepted in practice. For example, Hersh writes “what
mathematicians at large sanction and accept is correct” (2014, 149). We
should also note that:

(i) some past mathematical communities have been led astray and
accepted unrigorous (and incorrect) arguments – e.g., the
community working with infinitesimals when they were
introduced, the Italian school of algebraic geometry;

---

15 See (De Toffoli and Fontanari Manuscript).
16 A similar distinction is also drawn by another leading exponent of the school; Enriques
talks of “small-scale logic” and “large-scale logic.” See (De Toffoli and Fontanari
forthcoming).
(ii) some contemporary mathematical communities accept arguments that other communities deem as unrigorous – the Japanese community of mathematicians around Mochizuki;

(iii) the tradition of theoretical physicists has often been to accept mathematical arguments that are unrigorous by the standards of professional mathematicians,\(^\text{17}\) and this tradition continues to the present day – for example, in the work of Witten, whose results have been found immensely suggestive by mathematicians (he is the sole physicist ever to win the Fields Medal) but requiring further work to raise the arguments from physicists’ to mathematicians’ standards of acceptability.\(^\text{18}\)

4.4. What is, if anything, complete rigor?

Paseau (2016) argued that complete rigor is not, per se, epistemologically desirable. He assumed that an argument is completely rigorous if it is atomized. That is if it is composed of the smallest possible inferences. However, this assumption is highly controversial. One could argue that formal arguments are completely rigorous even if they include inferences that are not atomic.

4.5. Does rigor come in degrees?

Rigor can be used as an absolute concept – like in “genuine proofs are rigorous.” However, it is common to think of rigor as a matter of degrees. Jacobi once wrote in a letter to von Humboldt:

> Only Dirichlet, not I, not Cauchy, not Gauss, knows what a perfectly rigorous proof is, but we learn it only from him. When Gauss says he has proved something, I think it very likely; when Cauchy says it,

\(^{17}\) This tradition is long standing. By way of example, consider the use of unrigorous mathematical notions in Galilei’s study of the lever – contrasted with the (proto-)rigor treatment of the same problems by the ancients (Fano and Pietrini 2019).

\(^{18}\) One essential ingredient of the work for which Witten was awarded the Fields Medal in 1990 is the use of path integrals, which are generally recognized to lack mathematical rigor. In the words of Atiyah: “[Witten] has made a profound impact on contemporary mathematics. In his hands physics is once again providing a rich source of inspiration and insight in mathematics. Of course physical insight does not always lead to immediately rigorous mathematical proofs but it frequently leads one in the right direction, and technically correct proofs can then hopefully be found. This is the case with Witten's work” (1991, 34).
it is a fifty-fifty bet; when Dirichlet says it, it is certain; I prefer not
go into these delicate matters. Quoted in (Laugwitz 1999, 63).\textsuperscript{19}

One could think that a formal proof is always more rigorous than an
informal proof. But this is far from obvious. In fact, if rigor is supposed to
be what convinces of the existence of a formal proof, the best way to
convince mathematicians that a formal proof exists may not be by
exhibiting one.

4.6. What is rigor in intellectual endeavors other than mathematics?

Rigor can be seen as an intellectual virtue beyond mathematics. A
philosophical argument, for example, is rigorous when it is scrupulous.
Outside mathematics, rigor is, however, a much vaguer concept. A rigorous
argument can be shared among relevant experts. It is the kind of thing on
which a reasonable subject with the appropriate background training would
base a justified belief.

5. Conclusion

To sum up, we suggested that the problem of rigor as it faces us today
concerns the relationship between the logicians’ and the mathematicians’
conception of proof. More precisely, it arises from their different answers
to the question of how a rigorous proof manages to guarantee that the
premises invoked imply the conclusion reached. Although the origins of the
problem of rigor are to be traced already in ancient mathematics, its terms
became well-defined only with the development of modern logic, in the first
half of the nineteenth century. Still, given the contemporary description of
the problem, we can fruitfully apply it to various periods in the history of
mathematics.

The standard view cashes out rigor in terms of formalizability.
Notwithstanding its name, we showed that rather than a single view, the
standard view is a family of views. Although according to all the members
of the family, the rigor of an informal proof can be understood in terms of
formal proofs, they differ substantially on the details. Failing to
differentiate between them has led, in our opinion, to misunderstandings in
the literature. We then proposed a tentative list of choice-points (and

\textsuperscript{19}Thanks to Jeremy Avigad for pointing us to this quote.
potential answers) as a first attempt to differentiate the various specific positions belonging to the family of the standard view.

Lastly, we considered some of the ramifications of our general problem. A satisfactory account of rigor will in fact provide specific answers to a wide gamut of questions such as what the function of rigor is and whether rigor is a graded or an absolute notion.

References

---. Manuscript. “Recalcitrant Disagreement in Mathematics: an Endless and Depressing Controversy in the History of Italian Algebraic Geometry.”


sito web sono “no copyright”, nel senso che possono essere riprodotti, modificati, distribuiti, trasmessi, ripubblicati o in altro modo utilizzati, in tutto o in parte, senza il preventivo consenso di APhEx.it, a condizione che tali utilizzazioni avvengano per finalità di uso personale, studio, ricerca o comunque non commerciali e che sia citata la fonte attraverso la seguente dicitura, impressa in caratteri ben visibili: “www.aphex.it”. Ove i materiali, dati o informazioni siano utilizzati in forma digitale, la citazione della fonte dovrà essere effettuata in modo da consentire un collegamento ipertestuale (link) alla home page www.aphex.it o alla pagina dalla quale i materiali, dati o informazioni sono tratti. In ogni caso, dell’avvenuta riproduzione, in forma analogica o digitale, dei materiali tratti da www.aphex.it dovrà essere data tempestiva comunicazione al seguente indirizzo (redazione@aphex.it), allegando, laddove possibile, copia elettronica dell’articolo in cui i materiali sono stati riprodotti.