What Is the Epistemological Status of Computer-Assisted Proofs? An Empirically-Informed Approach[1]

Paul Hasselkuß
Heinrich Heine University Düsseldorf

Accepted manuscript. Please cite the published version.

ABSTRACT In the last few decades, beginning with Appel and Haken's proof of the four-color theorem, philosophers have been interested in the epistemological status of computer-assisted proofs. There are two opposing views. Critics point to mathematicians who criticize Appel and Haken's proofs as a verification that falls short of a real proof (e.g. Rota, 1997). They further argue that computer-assisted proofs are epistemically lacking because they cannot be surveyed or involve a specific risk of error (e.g. Tymoczko, 1979; Resnik, 1999). On the other hand, supporters point out that Appel and Haken's proof was accepted by those mathematicians who were actively working on the four-color theorem (MacKenzie, 1999). Furthermore, they try to refute the epistemic arguments of the critics (e.g. McEvoy, 2022). The result is a stalemate between critics and supporters. In this paper, I argue that this stalemate can be resolved by empirical data. I report the results of an analysis of all 2.6 million preprints uploaded to the arXiv between 1986 and 2024, conducted to find out exactly how many computer-assisted proofs are published, and how their number changes over time. The results show that there is a small but not insignificant number of preprints reporting on computer-assisted proofs. More importantly, their number has been increasing at an accelerating rate. This suggests that the epistemic concerns of the critics may be somewhat exaggerated.

1. Introduction

In the last few decades there has been a continuous interest in computer-assisted proofs. Beginning with Appel and Haken's breakthrough proof of the four-color theorem (Appel & Haken, 1977a; Appel et al., 1977) and

[1] Many thanks to the participants of the 2024 Masterclass in the Philosophy of Mathematical Practice at Vrije Universiteit Brussel and the Philosophy Master's and Doctoral Students Seminar at Heinrich Heine University, as well as to Deborah Kant and to two anonymous reviewers for their helpful comments on earlier versions of this paper.

what is taken to be a mixed reception in parts of the mathematical community, philosophers have been interested in two related questions: Are computer assisted-proofs genuine proofs, or do they merely verify their conclusions (e.g. Rota, 1997)? Moreover, are there any implications for the epistemology of mathematics, i.e. is the justification provided by computer-assisted proofs *a posteriori* because it is partly established by external means (e.g. Tymoczko, 1979)?

Answers tend to fall into one of two opposing camps, *critics* and *supporters*. Critics point to the lukewarm, at best, reception of computer-assisted proofs by the mathematical community. In their view, mathematicians who oppose computer-assisted proofs correctly recognize that these proofs are epistemologically lacking: they do not provide the kind of secure and indubitable justification that is normally associated with mathematical proofs (i.e. computer-assisted proofs are *a posteriori*). On the other hand, supporters argue that the critics' epistemological arguments are flawed in that they fail to distinguish between computer-assisted and ordinary, i.e. non-computer-assisted, proofs (i.e. computer-assisted proofs are *a priori* if ordinary proofs are).[2] Moreover, sociohistorical treatments of the Appel and Haken proof and its context may be taken to indicate that the proof was accepted by relevant parts of the mathematical community, namely by those mathematicians working on the four-color theorem (MacKenzie, 1999). Thus, while critics build on what they believe to be a negative response from the mathematical community, supporters can try to undermine that foundation by pointing to other voices within the community.

In this chapter, I'll argue that some progress can be made by taking an empirically informed view at the issue. Critics and supporters (though the latter perhaps to a lesser extent) rely on claims about mathematical practice, namely about what mathematicians think about computer-assisted proofs. These claims are often supported by pointing to the mathematical community. But *both* critics and supporters can find mathematicians whose statements seem to support their views. Who

[2] Note that there are also critics who avoid expressing their criticism in terms of the *a priori/a posteriori* distinction. I assume that such criticism is either *constructive* or *destructive*. Constructive criticism does not object to computer-assisted proofs, but points out flaws that could be improved (e.g. Thurston, 1994, p. 162). Destructive criticism objects to the usage of computer-assisted proofs altogether (e.g. Rota, 1997, p. 186). I am taking it that destructive criticism, ultimately, does make for an epistemic distinction between ordinary and computer-assisted proofs that can be voiced in terms of the *a priori/a posteriori* distinction.

should we trust? What do mathematicians really think about computer-assisted proofs? One answer, I'll argue, can be found through empirical research. I'll report on a bibliometric study that examines how common computer-assisted proofs are and whether their number changes over time by analyzing the 2.6 million preprints submitted to the arXiv between 1986 and 2024. The results address the above question indirectly: the number of computer-assisted proofs, as well as a possible decrease or increase in their number over time, is certainly indicative of the community's attitude towards them. In fact, my results suggest that some of the critics' claims may have been exaggerated.

In the next section, I'll give a more detailed overview of the arguments put forward by both, critics and supporters, as well as their respective views of the mathematical community's reception. Section 3 puts these views to the empirical test. I'll outline my approach and report the results. Finally, section 4 discusses the implications of these results in light of the arguments of both critics and supporters.

2. Critics vs. supporters

In this section, I'll examine the arguments offered by both critics and supporters. In particular, it will be interesting to see their respective commitments about what mathematicians think about computer-assisted proofs. While critics emphasize a negative attitude, supporters point out that an important group of experts has no objection to computer-assisted proofs. Historically, the debate is anchored in Appel and Haken's proof of the four-color theorem so I'll start there.

The four-color theorem states, in plain terms, that any planar map can be colored with four colors such that countries with a common boundary segment are given different colors. This can be translated to graph theory. Each country is represented by a vertex, and boundary segments between two countries are represented as connections between the corresponding vertices. So:

4CT Every plane graph has a 4-coloring.

The conjecture has been discussed since the middle of the 19th century, but a correct proof was not found until 1977. While this proof, by Appel and Haken, relies on a number of technical concepts that have been established through ordinary mathematical work (i.e. without the use of computers), the Appel and Haken proof relies on a computer to do some parts that are practically impossible for a human mathematician to do.[3]

What did Appel and Haken do? A detailed account would be out of place here, so I'll just give a brief, non-technical overview.[4] The overall strategy is to prove that there is no counterexample to 4CT, i.e., that there is no plane graph that has a coloring of five or more. To do this, the proof relies on two technical concepts: *unavoidable sets* and *reducible configurations*. An unavoidable set is a set of sub-graphs such that every possible plane graph must include at least one of its members somewhere. A reducible configuration is a sub-graph which, when found within a larger graph, renders the graph four-colorable due to certain structural properties. While there are many unavoidable sets and many reducible configurations, to prove 4CT, it suffices to show that there is a single unavoidable set that consists entirely of reducible configurations. This proves that there is no plane graph with a coloring equal to or greater than five.

Unavoidable sets can be found by a (complex) algorithm. The algorithm has been implemented and run computationally, but the published version of the proof contains an ordinary proof of the algorithm's correctness (Appel & Haken, 1977a).[5] Thus the correctness of the algorithm "can be checked by hand in a couple of months" (Appel & Haken, 1977b, p. 121). Reducibility is arguably more problematic. Checking whether a sub-graph is reducible ultimately amounts to a brute-force check of its individual features. While this is already time-consuming for smaller graphs, the graphs of the proof can have an outer ring size of 15 vertices – i.e. they are quite large – and, given the sheer number of graphs that need to be checked, "it would be virtually impossible to verify the reduction computations" (Appel & Haken, 1977b, p. 121) manually. Accordingly, the check for reducibility is implemented computationally.

---

[3] It is worth noting there were a number of unsuccessful attempts before the proof was found in 1977. The most famous of these, by Alfred Kempe in 1880, was not discovered to be faulty until 11 years later. A similar event took place shortly before Appel and Haken's proof, when Yoshio Shimamoto's proof attempt was refuted. Overall, these developments seem to have influenced a skeptical attitude towards proofs of the 4CT, and even towards the 4CT itself.

[4] For a detailed, but accessible outline of the proof, see Appel and Haken (1977b). The canonical source is Appel and Haken (1989), for the socio-historical context see MacKenzie (1999).

[5] This is the so-called 'discharging' algorithm, originally due to Heesch (1969).

The result of this procedure is an unavoidable set of 1.936 reducible configurations (the number could later be reduced to 1.405 by eliminating redundancies and simplifying of the argument, cf. Appel & Haken, 1989) and thus a proof of the 4CT.

To conclude the historical exposition, it is worth mentioning that Robertson et al. (1997) and Gonthier (2008) have published a new proof of the 4CT. This new proof is based on the same concepts – it still searches for an unavoidable set of reducible configurations – and is still computer-assisted, but it improves on the Appel and Haken proof by greatly simplifying the algorithm used to identify unavoidable sets. Moreover, the computational parts were implemented in Coq, a standardized proof assistant, whereas Appel, Haken, and their collaborator John Koch had to implement these parts in assembly language and run them on bare metal.[6]

What are the philosophically interesting aspects of this story? Critics begin by citing the "mixed" (Wilson, 2014, p. 157) reaction to the proof, with mathematicians expressing discomfort with the reliance on a computer to perform the reducibility check. This is acknowledged to some extent even by Appel and Haken:

> Many mathematicians ... resist treating the computer as a standard mathematical tool. They feel that an argument is weak when all or part of it cannot be reviewed by hand computation. From this point of view the verification of results such as ours by independent computer programs is not as convincing as the checking of proofs by hand. (Appel & Haken, 1977b, p. 121)

To give a few examples, Appel and Haken's proof has been criticized (by fellow mathematicians!) for not giving "a satisfactory explanation why the theorem is true" (Stewart, 1981, p. 304), partly because the computational parts are impossible to grasp. What Appel and Haken have archived is at best "a computer *verification* of the four color conjecture" (Rota, 1997, p. 186, my emphasis). The proof is not regarded as definitive and further verifications are needed. At worst, it amounts to "computer shenanigans [that] leave us intellectually unfulfilled." (Cohen, 1991, p. 328) The most telling example might be the following story of Haken, who recalls that when he visited another university to give a talk about Appel and his

[6] Assembly language is a low-level language in which each statement corresponds directly to a machine instruction. It requires manual management of low-level operations (e.g. memory addressing and instruction sequencing) and provides minimal abstraction and error checking. In contrast, Coq and other modern proof assistants use high-level languages that abstract and manage many low-level operations internally.

proof, a senior mathematician from that university actively tried to prevent him from engaging with local graduate students, apparently to protect them from the bad influence of his computational methodology (MacKenzie, 1995, p. 41).

But the critics do not stop there. The next step is to turn some of these points into an epistemic argument, i.e. to argue that computer-assisted proofs differ in an epistemically relevant way from their non-computer-assisted counterparts. The idea seems to be that those mathematicians who object to the use of computer-assisted proofs are actually responding to an epistemic difference between ordinary proofs and their computational counterparts.[7] Often, the aim is to argue that computer-assisted proofs provide an *a posteriori* justification for their conclusions, whereas ordinary proof might be expected to provide an *a priori* justification for their conclusions.[8]

To spell out the difference precisely, Tymoczko (1979, pp. 59-60) has coined the notion of surveyability:

S A proof is surveyable if and only if it can be completely looked over, reviewed, and verified by a rational agent.

Surveyability is of interest because it explains why a mathematician who reads a proof becomes convinced of the correctness of the conclusion. The proof leaves no room for doubt, i.e., if a mathematician is able to survey the proof and finds it to be valid, she must also accept its conclusion. Nothing about how the proof came about, or how it was found, is necessary to accept its conclusion.[9] The next step is to argue that Appel and Haken's proof is not surveyable because the reducibility part cannot be checked by hand (Tymoczko, 1979, p. 70).

[7] That is, it is assumed that the mathematical communities' reactions provide strong evidence for the epistemic quality of computer-assisted proofs. I will return to this point below.

[8] To complicate matters, Tymoczko (1979) and Kitcher (1984) claim that the justifications provided by ordinary proofs are also *a posteriori*. For them, the case of computer-assisted proofs is of interest because it brings to light what an otherwise overlooked problem. For the purposes of this paper, however, it is sufficient to focus on computer-assisted proofs.

[9] Recently it has been suggested to replace 'surveyability' with the more general notion of 'transferability', where a proof is transferable if and only if the sequence of propositions itself constitutes the proof (De Toffoli, 2021, pp. 9-11). While this may be a fruitful idea, for the context of this chapter the already established notion of surveyability will suffice.

The (un)surveyability has direct implications for the *a priori/a posteriori* distinction. Some authors argue that unsurveyable proofs do not constitute *a priori* justifications. Traditionally, *a priori* justifications have been characterized as indefeasible. This supposed indefeasibility is incompatible with justifications involving perception or recollections from memory: both are fallible and therefore defeasible (De Toffoli, 2021, p. 10). When applied to mathematical proofs, the consequences are clear. If a proof is surveyable, a skilled mathematician can easily follow each of its deductive steps and thereby gain an indefeasible justification for its conclusion. But if a proof cannot be fully surveyed, then any justification a mathematician might gain from the proof will be defeasible, because the mathematician cannot follow each of the proof's deductive steps, but has to accept the computer's output at some point. In counterfactual scenarios, contrary evidence – even empirical evidence – could undermine whatever justification the mathematician gained from the computer-assisted proof (Kitcher, 1984, p. 46).[10]

A second argument used by critics is also based on the lack of surveyability. A mathematician who reads the proof and accepts 4CT on the basis of her reading cannot in principle rule out that the computational parts of the proof contain programming errors or hardware failures (Resnik, 1999, pp. 150-153).[11] If the mathematician still believes that 4CT is correct, her belief is based not only on what she has read, but also on the (empirical) assumption that there were no programming errors or hardware failures, i.e. on a very general assumption about computers, perhaps to the extent that if such failures had occurred, they would not have produced the output necessary for the proof. The important point here is that this assumption goes beyond what the proof itself offers, and, so the critics argue, an analogous assumption is not needed in ordinary proofs. These can be surveyed, and therefore verified, by a (skilled) agent without having to think about how the proof was produced. Again, the implications for the *a priori/a posteriori* distinction are obvious. If accepting the result of a computer-assisted proof involves an empirical claim about the reliability of

---

[10] The same argument has been made against ordinary proofs which are very long (Kitcher, 1984, pp. 40-46). Within the framework outlined above, a mathematician conducting a long proof must memorize and later recall immediate steps. Since our memory is fallible, the justification of the proof is *a posteriori*. Be that as it may, to criticize the argument it suffices to show that *a priori* justifications are not indefeasible. I will return to this point below.

[11] Appel and Haken's proof was thought to contain such errors, although none could be found (MacKenzie, 1995, p. 42).

computers, an agent's belief in the conclusion of the proof can hardly be assumed to be justified *a priori*.

Supporters of computer-assisted proofs who wish to respond to these challenges seem to have two strategies at their disposal. One is to respond to the claim that the reception of computer-assisted proofs by the mathematical community has been overwhelmingly negative. A second strategy is to refute the epistemological challenges. I'll now discuss both strategies.

Modern mathematics is a highly specialized field. Mathematicians working in one sub-field often have great difficulty following the work of their colleagues working in another sub-field, let alone those working in another field (Thurston, 1994, p. 3). With this in mind, one may wonder how much weight should be given to the statements quoted by the critics. Most of the mathematicians who were quoted, although established scholars, have not been involved in recent graph theory or in the developments that led to Appel and Haken's proof. One may wonder what the mathematicians who have been involved in these communities think about computer-assisted proofs, and about Appel and Haken's proof in particular.

W. T. Tutte, a leading graph theorist who had been critical of earlier computational attempts to prove the 4CT, is known to have publicly touted Appel and Haken's proof as correct (Kolata, 1976). In addition, the core group of mathematicians who were actively working on the 4CT using methods similar to those of Appel and Haken – namely Frank Allaire, E.R. Swart, and Frank Bernhart – reacted quite enthusiastically to what they believed to be a successful proof:

> I can only tell you that we were unanimous in our view that [Appel and Haken had] done it. Unanimous. (Sward in an interview with A.J. Dale, quoted after MacKenzie, 1995, p. 42; cf. also Swart, 1980)

In light of these voices, supporters can argue that the critic's assessment of the mathematical community's stance is biased. While some mathematicians have objected to computer-assisted proofs, and in particular to the proof of the 4CT, the relevant experts have not. On the contrary, they have praised Appel and Haken's methodology and refuted the criticism of their colleagues (e.g. in Swart, 1980).[12]

The second strategy that the supporters can resort to is to respond to the epistemological arguments of the critics. More specifically, supporters argue that the notion of surveyability is epistemically negligible. McEvoy (2007) has argued that the very same proof can be represented differently, e.g. by using different notations. While one representation may not be surveyable, another representation of the very same proof may be. Thus one can turn an unsurveyable proof into a surveyable one by translating it from one representation into another. Given this possibility, McEvoy argues, surveyability does not seem to have any effect on the *a priori/a posteriori* distinction, contrary to what the critics claim. While a change in representation may transform a non-surveyable proof into a surveyable one, the change is unlikely to transform an *a posteriori* proof into an *a priori* one. This distinction does concern the proof and its justification, not its presentation. So the only viable explanation is that the proof (or rather the justification it provides) was *a priori* all along.[13]

Another response to the challenge focuses on the alleged connection between the possibility of doubt and the *a priori/a posteriori* distinction. Kitcher (1984), voicing the critical position, argued that the justification provided by computer-assisted proofs is defeasible because the proofs cannot be fully surveyed. Consequently, if *a priori* justifications are said to be indefeasible, the justification provided by computer-assisted proofs is *a posteriori*. However, responses from the epistemological literature challenge this conclusion. It has been argued that *a priori* justifications are in fact defeasible and can be undermined by both non-empirical and empirical evidence (Casullo, 1988).

A similar treatment can be given to Resnik's (1999, pp. 150-153) concern about implicit empirical assumptions. While Resnik is right about the presence of such assumptions, the same assumptions are also present in the case of ordinary proofs: "flaws in the computer implementation ... are nothing other than errors of logic, no different ... in proofs that have nothing to do with computer" (Swart, 1980, p. 703). While accepting the

---

[12] Historically, it is interesting to note that other mathematicians, including Allaire, Steward and Bernard, were working on their own computer-assisted proof of the 4CT (Wilson, 2014, pp. 150-151). Heesch had (unsuccessfully) attempted a computer-assisted proof before. In this sense, the option of attempting a computer-assisted proof may have been on the table for some time before Appel and Haken succeeded.

[13] See McEvoy (2022) for further responses to the critics.

result of a computer-assisted proof may involve some kind of trust in the inner workings of the computer, in practice a similar kind of trust is invoked when accepting the result of an ordinary proof (of a certain complexity and length).[14] In any case, it is argued that this aspect has no implications on the *a priori/a posteriori* distinction.

Where do we stand? So far I have painted a rather mixed picture. On the one hand, critics hope to argue that computer-assisted proofs are epistemically lacking, i.e. that they provide *a posteriori* justification where ordinary proofs might be expected to provide *a priori* justification. The argument is based on the notion of surveyability. It is argued that computer-assisted proofs are not surveyable, whereas ordinary proofs are. Supporters, in turn, argue against this verdict. While they agree that computer-assisted proofs may not be perfectly surveyable, they argue that this has no implications for their epistemological status.

Both positions have their roots in mathematical practice. Critics cite mathematicians who actively oppose the use of computer-assisted proofs to argue that there is something fundamentally wrong with these proofs. To explain what exactly is wrong, they appeal to the *a priori/a posteriori* distinction. In contrast, supporters cite mathematicians who are more open to computational methods. They argue that, contrary to the critics' portrayal, mathematicians are not universally opposed to computational methods: while some reject their use, others embrace them. While this divergence may still call for an explanation, it challenges the idea that there is something *fundamentally wrong* with computer-assisted proofs.

Both critics and supporters presuppose that the mathematical community acceptance of a proof or method is a good indicator of its epistemically quality. This presupposed *epistemic link* can be spelled out as follows:

EL The mathematical community's acceptance or rejection of a method or proof provides strong evidence of its epistemic quality – either positively or negatively.

While this assumption is common at least in the epistemological strand of the philosophy of mathematical practice (Carter, 2019, pp. 16-22), the appeal to (EL) is not entirely unproblematic. Community consensus may also reflect sociological or cultural trends, rather than purely epistemological considerations. Also, different parts of the community

[14] Swart adds that, in comparison, he trusts the parts of Appel and Haken's proof that were done by a computer more than the parts that were done by hand.

may hold different views. When the supporter argues against the critic that relevant parts of the mathematical community have accepted Appel and Haken's proof, she is pointing out that the critic has mistakenly appealed to (EL), i.e. the critic mistakenly assumes that the whole community objects to Appel and Haken's proof, whereas only some parts of the community have objected. However, given that critics and supporters share their commitment to (EL), it seems unproblematic to adopt it for the present discussion.[15]

This leaves us at a stalemate: if critics rely on those mathematicians who oppose computational methods and supporters rely on those who embrace them, who is correct? What do mathematicians really think about computer-assisted proofs? And what is their epistemological status? In the next section, I argue that this question ultimately requires empirical investigation.

3. Investigating the issue empirically

In this section, I'll argue that both critics and supporters not only use claims about the mathematical community's reception of computer-assisted proofs to support their respective views, but also make implicit assumptions about how the mathematical community would be expected to behave or think if their respective views were correct. And these assumptions can (and arguably should, cf. Aberdein & Inglis, 2019) be tested empirically. Below, I'll describe a methodological approach to testing these assumptions by analyzing preprints submitted to the arXiv, and I'll report the results. But first, let me make the implicit assumptions explicit.

The critic points to the statements of mathematicians who react negatively to computer-assisted proofs to support her epistemological points, i.e. she argues that the mathematicians' discomfort is strong evidence that something is epistemically wrong with computer-assisted proofs (i.e. she commits herself to (EL) as outlined above). Next, the critic points to the *a priori/a posteriori* distinction to answer what is epistemically wrong with

[15] Note that this does not imply that any critical or supporting position must be based on (EL). For instance, a critic might argue that computer-assisted proofs are epistemically flawed based on purely epistemic reasons, irrespective of the community's stance. However, as far as I have reviewed the existing literature above, this is not how the debate has unfolded. Critics begin with observations about opposition to computer-assisted proofs within the mathematical community and infer epistemological consequences from there.

computer-assisted proofs. While the critic provides some evidence for the starting claim by citing mathematicians who have reacted negatively, the critic is nevertheless committed to the broader claim that the mathematical community *as a whole* reacts negatively. Only this broader claim can be used to invoke (EL). And this broader claim can be tested empirically.

One metric to track acceptance or rejection is the number of computer-assisted proofs, or rather the number of papers including computer-assisted proofs, that are published. If computer-assisted proofs are not accepted by the mathematical community, one would expect that, apart from perhaps a few mavericks, mathematicians would not want to write and publish computer-assisted proofs. Thus, the critic is committed to the following, arguably rather broad, estimation claim:

C1 Computer-assisted proofs are unpopular and rarely found within mathematical practice.

More specifically, it might be interesting to consider whether popularity is expected to change over time. Appel and Haken may have been pioneers in their field, but at the time of their work, computer-assisted proofs were not exactly popular.[16] Since then, however, mathematicians may have become more accustomed to computational methods. Appel and Haken speculate that their proof faced opposition, particularly from those mathematicians "educated before the development of high-speed computers" (Appel & Haken, 1977b, p. 121). With the huge advances made in the development of (not only) high-speed computers in the last 50 years, and the increasing popularity of informatics and computer science among mathematics students, can we expect that the number of computer-assisted proofs will increase?

The critic, I think, would give a negative answer. If computer-assisted proofs are not accepted (for epistemic reasons, as the critic will want to claim), and if mathematicians therefore shy away from engaging with them (as (EL) suggests), then the critic would expect that this will *not* change. The critic would thus be committed to the following relational claim (considering the relation between different points in time and the number of computer-assisted proofs):

C2 Over time, the number of computer-assisted proofs remains (more or less) unchanged.

---

[16] To my mind, Appel and Haken's proof was the first computer-assisted proof published.

Together with C1, C2 implies that the number of computer-assisted proofs is and remains low.

Let us now look at the supporter's view of the matter. In the supporter's view, the mathematical community *as a whole* is not reacting negatively towards computer-assisted proof. Although some mathematicians object to their usage, they are the exception rather than the norm. Most mathematicians may be indifferent, while some may even endorse computational methods. In any case, on the supporter's view, there is not enough evidence to invoke (EL) to argue that there is a fundamental epistemic problem.[17] Computer-assisted proofs may not be perfect, but the supporter holds that there is nothing epistemically wrong with them, simply because there is no universal rejection. Accordingly, the supporter might express the following expectation:

S1 Computer-assisted proofs are found in mathematical practice.

Note that the supporter does not have to claim that computer-assisted proofs are particularly popular, or that the number of proofs found exceeds some (fairly high) threshold. What might be expected is that their number is greater than the critic would expect, but there is no need (or way) to quantify this assumption. Perhaps it is fair to say that, on the supporter's view, computational methods are expected to be endorsed by wider parts of the mathematical community, and not just by a few mavericks.

What is the supporter's corollary to C2? Given the tremendous advances in the development (and widespread availability) of fast computers and the increasing popularity of computer science, the supporter might adopt the following:

S2 Over time, the number of computer-assisted proofs is rising.

We now have four empirical claims about mathematical practice, namely about the number of computer-assisted proofs that can be found. These claims follow from the respective philosophical positions of critics and supporters, and their reliance on (EL). When these claims are tested empirically, the corresponding philosophical position is also tested.[18]

---

[17] Needless to say, there is not enough evidence to invoke (EL) to argue that computer-assisted proofs are epistemically good, either.

[18] Note that C2 and S2 are incompatible. As the results will show, the crucial finding concerns how the number of computer-assisted proofs changes over time, one could simplify things and focus exclusively on C2.

Next, I'll describe a method for doing just that.

3.1 Method

From the way claims C1, C2, S1, and S2 are worded, it should be obvious that I am interested in the number of computer-assisted proofs and how this number changes over time. One method of generating the relevant data is bibliometric analysis.[19] The plan is as follows: first, one needs a way to access a sufficiently large bibliographic dataset that is representative of the mathematical practice and the community one wants to study. Then one needs to devise a way to extract computer-assisted proofs from what is available in the dataset. Finally, one counts the number of computer-assisted proofs and checks if and how their number has changed over time. The results can be compared with the expectations set out in C1, C2, S1, and S2. I'll now describe each of these steps.

Given the need for a representative dataset, it is common practice to rely on the arXiv (e.g. in Mejía-Ramos et al., 2019 and Tanswell & Inglis, 2024). The arXiv is a repository for preprints of academic articles in fields such as mathematics, computer science, physics, quantitative biology, statistics, electrical engineering and systems science. It is the primary repositories that mathematicians around the world use to share their work (McKinney, 2011). It is therefore reasonable to assume that by using data from the arXiv, one will get a dataset that is representative of current mathematical practice and community. Another advantage of the arXiv is that it is open access. Preprints and their metadata can be freely viewed and downloaded without technical or legal hurdles. For the purposes of this chapter, I have used the so-called 'arXiv dataset', which contains the metadata of all 2.6 million articles currently on the arXiv and is available for download on Kaggle (arXiv.org, 2025).

Having a representative dataset, the next step is to extract the preprints that contain computer-assisted proofs. To do this, I have devised a list of keywords, each of which is assumed to be indicative of a computer-assisted proof. The idea is that any preprint that contains at least one of these keywords is likely to report on a computer-assisted proof. In technical jargon, this amounts to a Boolean information retrieval without ranking (Manning et al., 2008, pp. 1-17). The arXiv dataset forms the collection (or corpus) from which information is retrieved, while the metadata of

[19] To be clear, bibliometrics is certainly not the only way, but only one of many empirical ways to study the question. In fact, it may by advisable to use different methods to study the same question (Löwe & Kerkhove, 2019).

each preprint are the (so-called) documents within the corpus. The Boolean condition consists of the list of keywords given below.

Keywords

ACL2, Agda, Coq, HOL, Idris, Isabelle, Lean, Metamath, Mizar, computer-assisted, computer assisted

The keywords include the terms 'computer-assisted' and 'computer assisted', the latter to catch a common typo. It also includes the names of all proof assistants listed in the Wikipedia article on proof assistants, provided that there is at least one article in the arXiv, reporting on a computer-assisted proof that includes the name.[20]

Next, all entries in the dataset were matched against the list of keywords. For practical reasons, matching was limited to both titles and abstracts. It is assumed that if a paper contains a computer-assisted proof, rather than an ordinary one, this information will be given in either the title or the abstract. Initial results confirm that this tactical assumption is feasible.

It is a unique feature of the arXiv that both titles and abstracts often contain LaTeX code. Previous projects built around data extracted from the arXiv have devised ways to strip the LaTeX code in order to analyze the underlying text (Mejía-Ramos et al., 2019; Tanswell & Inglis, 2024). For this project, however, it quickly became clear that such processing was not necessary. Initial results showed that the keywords, if present, did not interfere with the LaTeX code. For the same reason, it was not necessary to lemmatize the text (i.e., reduce words to their base or dictionary form to facilitate corpus-based analysis), although lemmatization would probably be necessary if the title and abstract of matching papers were analyzed further. However, it was necessary to match the keywords in a case-insensitive-manner, i.e. an article was counted as a match if its title or abstract included one of the keywords, regardless of the case of the word. Initial results confirmed that this works, except for 'HOL' and 'Lean', both of which result in false positives when matched case-insensitively.[21] However, these could be remedied by case-sensitive matching of these two words.

---

[20] This was confirmed by entering the term in the arXiv online search and manually checking the resulting preprints.

[21] The term 'lean' is often used in its ordinary English sense, rather than referring to the proof assistant of the same name. 'HOL' is sometimes used to name a function within mathematical formulae.

Matching and statistical computations were performed using R (version 4.4.2; R, 2024), the code is available on GitHub.[22] Let me now report the results.

3.2 Results

The dataset contains 2.638.713 preprints submitted between 1986 and 2024. (Preprints newer than 2024 were removed before conducting the analysis.) A total of 2.652 matched the condition, i.e. they were classified as reporting on computer-assisted proofs because their title or abstract contained at least one of the keywords (0.1%).

The next step was to see if the number of matches per year changed. It was found that until the year 1999 there were almost no matches, but then the number of matches increased relatively quickly. In the year 2000 there

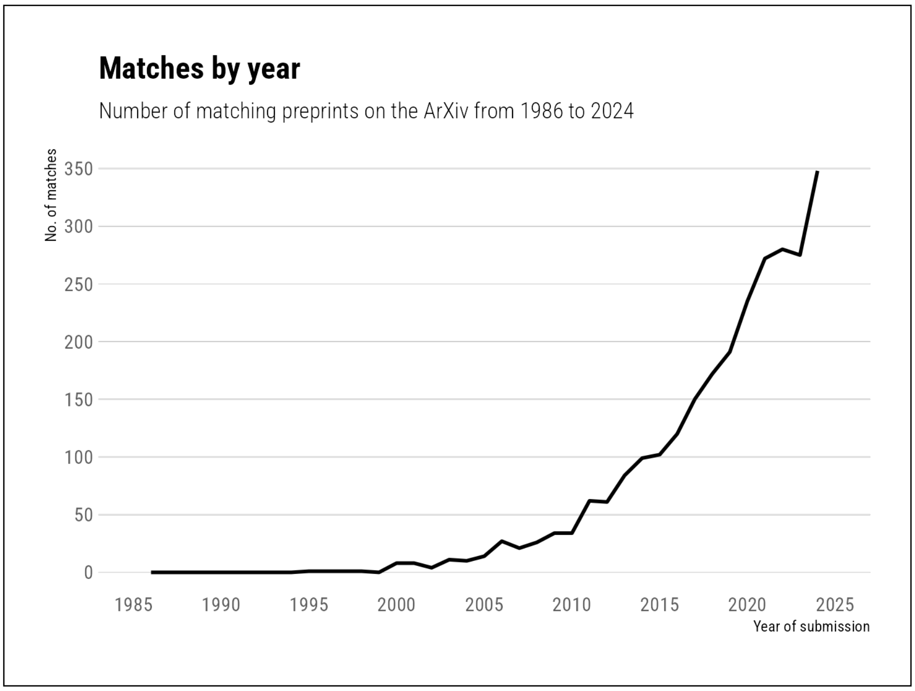

Fig. 1 Preprints matching at least one of the keywords and uploaded between 1986 and 2024. Each preprint is counted exactly once.

were 8 matches, while in the year 2024 there were 348 matches. Figure 1 shows the number of matches per year, the details are listed in Table 1 in the appendix.

In order to estimate how the number of matches grows over the years, a quadratic regression analysis was carried out, i.e. it was investigated

[22] https://github.com/paulHasselkuss/computer-assisted-proofs-arxiv

whether the growth rate itself changes over time. The model was statistically significant, $F(2,35)=542.1$, $p<.001$, and explained a significant proportion of the variance in the number of matches, $R^2=.967$ (adjusted $R^2=.967$). The first degree polynomial term was a significant predictor, $b=510.59$, $SE=17.98$, $t(35)=28.4$, $p<.001$. The second degree polynomial term also significantly predicted the number of matches, $b=299.60$, $SE=17.98$, $t(35)=16.66$, $p<.001$. This suggests that the number of matches is not just increasing steadily, but at an accelerating rate.

If we look at the number of preprints submitted to the arXiv per year, we can see that the number of preprints also seems to be increasing at an accelerating rate (see Table 1 in the appendix). Thus, it may be natural to assume that the increase in matches is merely due to the increase in the number of preprints submitted, i.e. while the *absolute* number of matches increases over the years, the proportion of matches *relative* to the number of preprints submitted may remain unchanged.

A logistic regression analysis was performed to access the relationship between year and the proportion of matches. The model was statistically significant, $\chi^2(1)=599.83$, $p<.001$, and showed a strong relationship between the year and the likelihood of a match. Year was a significant positive predictor of matches, $b=0.076$, $SE=0.0035$, $z=21.94$, $p<.001$. This indicates that the probability of a match *increases* by approximately 7.9% with each passing year, i.e. the increase in matches *exceeds* the increase in preprints submitted (contrary to the assumption outlined above).

So far, I have reported the results of looking at papers from all fields that can be submitted to the arXiv. This includes academic fields that are relevant to the philosophical question (such as mathematics and computer science), but also fields that are not (such as astrophysics or nonlinear science). To make the results more precise, in the next step I looked only at preprints that were assigned to relevant categories. However, this is not entirely unproblematic, as a single preprint is often assigned to several categories, i.e. a unique preprint can be assigned to both mathematics and astrophysics. Thus, preprints have to be counted once *for each* category to which they are assigned, and the resulting counts cannot be compared with the counts of the preprints reported above.

Looking at the number of matches per category, most matches are found in computer science (2.145 matches, 0.3% of 707.583 preprints within the category), followed by mathematics (695 matches, 0.1% of 675.005 preprints), and electrical engineering and systems science (170 matches, 0.18% of 97.045 preprints). Matches in the remaining categories sum to 271 (0.01% of 1.974.215 preprints). Figure 2 shows the number of matches per category and year, while details are given in Table 2 in the appendix.

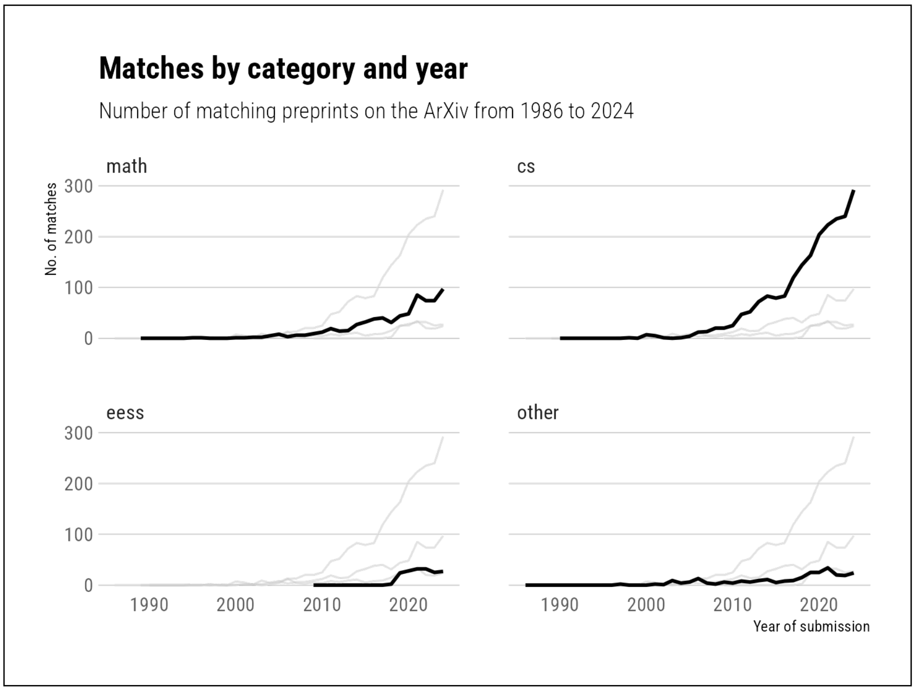

Fig. 2 Preprints matching at least one of the keywords and uploaded between 1986 and 2024, listed by category. As a unique preprint can have several categories, a unique preprint may be counted for several categories.

The categories mathematics and computer science were subjected to further scrutiny. For mathematics, a quadratic regression analysis was performed to estimate how the number of matches within the category increases over the years. The model was statistically significant, $F(2,33)=253.7$, $p<.001$, and explained a significant proportion of the variance in the number of matches, $R^2=.939$ (adjusted $R^2=.935$). The analysis revealed a significant positive linear effect of year on the number of matches, $b=134.26$, $SE=6.85$, $t(33)=19.6$, $p<.001$, and a significant positive quadratic effect, $b=76.03$, $SE=6.85$, $t(33)=11.10$, $p<.001$. This indicates that the number of matches within the mathematics category increased at an accelerating rate over time.

A logistic regression analysis was performed to investigate whether the increase in matches exceeded the increase in preprints submitted within the category. The model was statistically significant, $\chi^2(1)=144.14$, $p<.001$, indicating that the year is a strong predictor of the proportion of matches. The analysis revealed a significant positive effect of year on the odds of a match occurring, b=0.084, SE=0.008, z=11.0, p<.001. The results suggest that as the year increases, the probability of a match within mathematics *increases* by approximately 8.7% per year, i.e. the increase in matches *exceeds* the increase in preprints submitted.

For computer science, a quadratic regression analysis was performed to estimate how the number of matches within the category increases. The model was statistically significant, $F(2,32)=633.4$, $p<.001$, and explained a significant proportion of the variance in the number of matches, $R^2=.975$ (adjusted $R^2=.97$). The analysis revealed a significant positive linear effect of year on the number of matches, b=423.0, SE=13.91, $t(32)=30.91$, p<.001, and a significant positive quadratic effect, b=245.56, SE=13.91, $t(32)=17.65$, p<.001. These results suggest that the number of matches within the category of computer science has increased at an accelerating rate over time.

A logistic regression analysis was performed to study whether this effect exceeded the increase in submissions within the category. The model was statistically significant, $\chi^2(1)=79.18$, p<.001, indicating that year was a significant predictor of the proportion of matches. The analysis revealed a significant negative effect of year on the odds of a match occurring, $b=-0.041$, SE=0.004, $z=-9.35$, p<.001. The results suggest that as the year increases, the odds of a match *decrease* by approximately 4.05% per year, i.e., in contrast to mathematics, the increase in matches in computer science *is slightly less than* the increase in preprints submitted.

A final metric I'd like to report is the number of matches by keyword. Although it has no direct bearing on the philosophical issues, it is interesting (and perhaps unsurprising) that the term 'computer-assisted' has the most matches with 734 (27.68%), followed by the names of the proof assistants Coq with 607 (22.89%) and Isabelle with 380 matches (14.33%). The complete results are given in Table 3 in the appendix.[23]

[23] Caution should be exercised in drawing too much conclusions from these data. The present study only considered titles and abstracts. It is likely that many papers use the term 'computer-assisted' in their title or abstract but mention the software by name only in the text of the paper, or vice-versa.

4. Discussion

What conclusions can be drawn about the philosophical issue of computer-assisted proofs? Do these conclusions offer arguments in favor of the critic or the supporter? To answer these questions, I'll discuss the merits of each of the claims introduced at the beginning of the previous section in the light of the results. I'll conclude with some critical remarks.

C1 Computer-assisted proofs are unpopular and rarely found within mathematical practice.

The critic adopts C1 as a consequence of the claim that computer-assisted proofs are epistemically lacking, whereas ordinary proofs are not. Thus, the number of computer-assisted proofs is expected to be quite small. In order to assess whether this claim is supported by the empirical evidence, it is necessary to qualify what is meant by 'small'.

On a strong reading, 'small' could be taken to mean that there are only a handful of computer-assisted proofs (i.e. the method is only advanced by mavericks and not by serious mathematicians). This strong interpretation is not supported by the empirical results: there are a considerable number of computer-assisted proofs both across all categories (2.652 matches) and within the category mathematics (695 matches). Although these numbers are small compared to the number of non-matching preprints (0.1% for all categories and 0.1% for mathematics), they might be larger than the strong reading would suggest.[24]

However, on a weaker reading, C1 may be taken to mean that computer-assisted proofs are not advanced in the majority of published papers. The weaker reading is supported by the empirical findings. Computer-assisted proofs are found, but their number remains small. In both mathematics and computer science, the vast majority of preprints do not contain matches, and while there are still a significant number of matches, they are far too small to become the majority in the near future. So while the strong reading of C1 is not supported by the empirical evidence, the weak reading is.

[24] A first analysis also shows that the matching papers are *not* written by a small group of very productive authors. In total, the 2.652 matching papers have more than 5.000 unique authors. It should be noted, however, that while these numbers show a clear trend (i.e. a large number of individual authors), further empirical work would be needed to make these numbers precise. For example, the names as they appear in the arXiv dataset contain a larger number of mismatches (e.g. missing or differently spelt surnames) that would need to be addressed.

C2 Over time, the number of computer-assisted proofs remains (more or less) unchanged.

To develop C2, I considered the idea that because of the advances in high-speed computing over the last 50 years, and the widespread availability of sufficiently fast systems (you no longer need a supercomputer to prove 4CT, an old laptop will do), it might seem natural that the number of computer-assisted proofs would also increase. In my reconstruction, the critic rejects this assumption. Because computer-assisted proofs are thought to be epistemically deficient – because they amount to "shenanigans" (Cohen, 1991, p. 328) –, their number is not increasing, even though sufficiently fast computers become more accessible.

The empirical results clearly do not support C2. Looking at preprints across all categories, the number of matches has increased over the years. But it is not just increasing *steadily*, it is increasing at an *accelerating* rate. This suggests a rapidly growing trend in the number of computer-assisted proofs in recent years: while their absolute number remains low, more computer-assisted proofs are being published each year. Moreover, this effect outweighs the increasing number of preprints submitted, i.e. although the number of preprints submitted per year is also increasing, the proportion of computer-assisted proofs submitted per year is increasing even more.

A similar picture emerges when looking at preprints classified as mathematics and, with a caveat, those classified as computer science. For both, the number of preprints reporting on computer-assisted proofs is not only increasing steadily, but at an accelerating rate. For mathematics, this effect outweighs the increasing number of preprints submitted (i.e., the number of computer-assisted proofs categorized as mathematics is increasing faster than the number of all submitted papers categorized as mathematics). This is not the case for computer science. While the number of preprints categorized as computer science is increasing, the number of computer-assisted proofs categorized as computer science is not increasing at the same rate. Unfortunately, the available data do not provide an answer as to why this is the case. It could be that other areas of computer science are receiving more attention, or that preprints reporting on computer-assisted proofs are nowadays more likely to be assigned only to their primary category (e.g. mathematics or eess), rather than also being assigned to computer science.

In any case, C2 is not supported by the empirical data. On the contrary, the data suggest that the number of computer-assisted proofs is increasing at an accelerating rate.

S1 Computer-assisted proofs are found in mathematical practice.

Supporters defend S1 in response to C1, i.e. to express the expectation that computer-assisted proofs exist and are being advanced by mathematicians. As with C1, a distinction can be made between a strong and a weak reading. A strong reading would expect computer-assisted proofs to be found in a large number of publications (though perhaps not in the majority). A weaker reading would make no such assumption, but would simply expect to find a not insignificant, though perhaps not large, number of computer-assisted proofs.[25]

Clearly, the empirical results do not support the strong reading, as the number of matches was found to be relatively low (0.1% for all categories, 0.1% for mathematics, 0.3% for computer science). However, the results do support the weak reading. A total of 2.652 matches across all categories, and, although these figures cannot be directly compared because a single preprint can be assigned to several categories, 2.145 matches within computer science and 695 matches within mathematics, represent a small but significant proportion of the total.

S2 Over time, the number of computer-assisted proofs is rising.

The supporter's S2, like the critic's C2, formulates an expectation about how the number of computer-assisted proofs will change over the years. While the critic, in my reconstruction, denies that the number has changed significantly, the supporter, perhaps pointing to technological and societal advances over the last 50 years, assumes that the number of computer-assisted proofs is increasing.

Indeed, S2 is well supported by the empirical evidence. Looking at preprints across all categories, we see that the number of matches is not only *steadily* increasing, but is increasing at an *accelerating* rate. In fact, it has increased faster than the total number of preprints submitted, i.e. more and more of the preprints submitted are reporting on computer-assisted proofs. A similar conclusion was reached when looking at preprints in the category mathematics. Within mathematics, too, the number of

[25] I do not think that any supporter has seriously advanced the strong reading.

computer-assisted proofs is increasing faster than the total number of preprints submitted.

Where do we stand? What is the epistemological status of computer-assisted proofs? The empirical findings I have reported suggest that while computer-assisted proofs are not the primary strategy that mathematicians use when working on a problem – they are probably not the second or third either – they do seem to have their place in the mathematical community, at least as far as that community is reflected in the preprints submitted to the arXiv. While the number of computer-assisted proofs is relatively small, it is increasing and, more interestingly, increasing at an accelerating rate. This finding seems to suggest that the epistemological problems raised by the critic may have been somewhat exaggerated. If the critic bases their argument against computer-assisted proofs on (EL), claiming that they are rejected by the mathematical community as evidence of their negative epistemic quality – framed in terms of the *a priori/a posteriori* distinction – the data indicate that computer-assisted proofs are *no longer*, or perhaps never were, rejected by the community. This tips the scale towards the supporter: while not entirely without problems, computer-assisted proofs do not seem to be epistemically lacking.

I foresee at least four possible objections to this conclusion. First, one might object that the mathematical community is not epistemically well-behaved. That is, contrary to (EL), the mathematical community's acceptance or rejection of a method or proof does *not* provide evidence for its epistemic quality. On this view, the data I report would bear no relevance for the epistemic status of computer-assisted proofs.

I have already touched on this kind of reasoning in the discussion of (EL) above. As I noted there, both critics and supporters seem to agree on (EL) by pointing to mathematicians who reject or support computational methods. So I do not think that giving up (EL) is a plausible move in the context of the literature I have reviewed here. Moreover, one might ask why one should care for computer-assisted proofs if not because of (EL). If one does not assume that the mathematical community is epistemically well-behaved, why should one want to make an epistemic distinction between ordinary and computer-assisted proofs to begin with? Notions like surveyability are discussed because some mathematicians are concerned about computer-assisted proofs. Without (EL), why should one try to recast these in epistemic terms?

The second objection accepts (EL), but objects to applying corpus-based methods to study whether the mathematical community accepts or rejects a method or proof. It can be motivated as follows: acceptance or rejection of a particular method by the mathematical community ultimately comes down to whether individual mathematicians accept or reject proofs using that method. This is not a simple binary decision, but a complex and gradual process involving various factors. Corpus-based tools are thought to be too coarse-grained to adequately measure this complexity. The objection is therefore methodological. It assumes that there is a gap between what is measured – the frequency of words or the number of preprints containing those words – and the complex feelings of the agents that make up the mathematical community.

I think the objection is correct in identifying a potential gap between what is measured and scientific practice. But I also think it greatly exaggerates the gap, while underestimating what corpus-based tools are able to capture. First, note that there is always some sort of gap when applying empirical methods to a philosophical question. Even if one resorts to qualitative interviews, say, to study what mathematicians think about computer-assisted proofs, conducting and interpreting the interviews requires some sort of generalization to bridge the gap between what was measured (the individual responses) and the philosophical question (the mathematical practice). What matters is how this generalization is justified.[26] Second, studying mathematical publications and preprints certainly is a justified way to study mathematical practice. If only because publication practices are part of mathematical practice, so that by studying the former one can (within certain limits) make justified generalizations about the latter (cf. also Lean et al., 2023).[27]

The third objection I anticipate accepts (EL) and the use of corpus-based methods, but insists that the results I have reported above do not show enough to arrive at definitive conclusions about the epistemic status of computer-assisted proofs. There might be finer details in how the community, or individual mathematicians, think about computer-assisted

[26] If interviews are used, the selection of interviewees, the catalogue of questions and the coding of the responses must be justified. If corpus-based tools are used, one needs to justify why the corpus chosen is representative and why the keywords chosen are relevant for the underlying issue. As I have argued in section 3, the arXiv is representative of mathematical practice, and the keywords are indicative of computer-assisted proofs.

[27] This does not mean that other empirical methods should not also be used to study what mathematicians think about computer-assisted proofs. On the contrary, using different methods might be advisable to cover different aspects of the phenomenon.

proofs. Perhaps there are certain areas or contexts in which computer-assisted proofs are more likely to be accepted, but others in which they are not. There might also be a distinction between a computer-assisted proof of a new theorem (as was the case with 4CT), and formal verifications of theorems that have already been proven in traditional ways. Additionally, there is probably a historical dimension to these questions, in so far as early computer-assisted proofs, like Appel and Haken's proof of 4CT, are far more difficult to read than contemporary ones that are much more standardized.

In response, I do agree that it is highly plausible that such subtle differences exist. Studying them would require more fine-grained analyses than I have presented here. Corpus-based tools could still be helpful. For example, the 2.652 matching preprints I have identified could be studied further to learn more about their authors, and about the inner-mathematical fields in which computational methods are applied. However, I also want to point out that the question I have investigated here is far less complex than the objection suggests. What is the epistemological status of computer-assisted proofs? I have presented two views, critics and supporters, and argued that both are committed to empirical hypothesis about the number of publications reporting on computer-assisted proofs. I then tested these hypotheses and found little evidence to support critic's claim. This does not mean that computer-assisted proofs are without epistemic problems, but only that they do not seem to be rejected by the mathematical community in the way critic claims.[28]

The last objection I would like to consider is the relatively small numbers of matches. Even if one agrees with my reasoning up to this point, one might still claim that the number of matching preprints I have identified, that is 2.652 matches across all categories (0.1%), and 2.145 matches within computer science (0.3%) plus 695 in mathematics (0.1%),[29] is simply too low to claim that the mathematical community does not reject computer-assisted proofs.

---

[28] To take up the alleged distinction between computer-assisted proofs and formal verifications, the point is that the critic (as I have framed her) would want to insist that *both* are epistemically bad and therefore rejected by the mathematical community. My findings suggest that this is not the case.

[29] Note again that a single preprint can be assigned to more than one category. Therefore, the total number always tends to be lower than the sum of the numbers in the individual categories.

As I have argued above, I agree as far as the estimation claims C1 and S1 are concerned. The number of matches I found seems too low to pick one over the other. But I also found that the number of preprints reporting on computer-assisted proofs is increasing each year. Moreover, it is increasing at an accelerating rate, exceeding the increase in submissions (except in computer-science). This, as I have argued, is contrary to the assumption of the critics (cf. C2), whereas it is consistent with the assumption of the supporters (cf. S2). So although the absolute numbers are relatively low, I claim that their accelerating increase establishes that computer-assisted proofs are not rejected by the mathematical community.

5. Summary

What is the epistemological status of computer-assisted proofs? To answer this question, both critics and supporters rely on quotes from mathematicians who either criticize or support the use of computational methods. Critics argue that mathematicians who oppose computer-assisted proofs have a legitimate concern because computer-assisted proofs are epistemically lacking. Supporters counter by pointing out that relevant experts accept computer-assisted proofs and that the epistemic arguments of the critics can be rejected. This stalemate can be resolved, I have argued, by empirical studies. I have presented a study that analyses all preprints submitted to the arXiv from 1986 to 2024 to count exactly how many preprints reporting on computer-assisted proofs have been published, and to investigate how this number changes over time. The results show that there is a small but significant number of computer-assisted proofs on the arXiv and, more importantly, that their number is increasing at an *accelerating* rate. While more and more preprints are submitted to the arXiv each year, my results suggest that the proportion reporting on computer-assisted proofs is increasing even faster. Overall, this tips the scales in favor of the supporters: while mathematicians may not exactly like computer-assisted proofs, the epistemic concerns of the critics may have been somewhat exaggerated.

Zooming out, one might ask whether the stalemate is also due to a methodological difficulty. For example, Rota, a mathematician and strong critic of computational methods, has this to say:

> [Appel and Haken's] "proof" was the first verification of a major mathematical theorem by computer. Mathematicians have been ambivalent about such a verification. On the one hand, every mathematician professes to be satisfied to learn that the conjecture has been settled. On the other hand, the behavior of the community of mathematicians belies such a feeling of satisfaction. Indeed, if mathematicians had been satisfied with the computer verification of the four color conjecture, then no one would not have felt the need for further verifications. (Rota, 1997, p. 186)

Methodologically, Rota makes generalizations about the mathematical community and his fellow mathematicians to support his critique. While these may have some *prima facie* plausibility, given that Rota is an eminent mathematician who may have some privileged insights, even eminent mathematicians can err when generalizing about their discipline (Hanna & Larvor, 2020). Rota's view may be representative of a very particular community, but not so much of other communities.[30] But without empirical research, there is no way to know. Empirical research on mathematical practices and communities is needed to produce the evidence that provides viable input to philosophical theorizing (Buldt et al., 2008; Aberdein & Inglis, 2019). And, as perhaps in the case of computer-assisted proofs, if done properly it can prevent bias in one direction or another.

[30] As for Appel and Haken's proof, as mentioned above, the experts who were actively working on the 4CT at the time were by no means ambivalent, but accepted the proof as definitive.

References

Aberdein, A., & Inglis, M. (Eds.). (2019). Advances in Experimental Philosophy of Logic and Mathematics. Bloomsbury. https://doi.org/10.5040/9781350039049

Appel, K., & Haken, W. (1977a). Every planar map is four colorable. Part I: Discharging. *Illinois Journal of Mathematics*, *21*(3), 467. https://doi.org/10.1215/ijm/1256049011

Appel, K., Haken, W., & Koch, J. (1977). Every planar map is four colorable. Part II: Reducibility. *Illinois Journal of Mathematics*, *21*(3), 336. https://doi.org/10.1215/ijm/1256049012

Appel, K., & Haken, W. (1977b). The Solution of the Four-Color-Map Problem. *Scientific American*, *237*(4), 108–121. https://doi.org/10.2307/24953967

Appel, K., & Haken, W. (1989). *Every Planar Map is Four Colorable*. American Mathematical Society. https://doi.org/10.1090/conm/098

arXiv.org, S. (2025). *arXiv Dataset*. https://www.kaggle.com/dsv/7548853

Buldt, B., Löwe, B., & Müller, T. (2008). Towards a New Epistemology of Mathematics. *Erkenntnis*, *68*(3), 309–329. https://doi.org/10.1007/s10670-008-9101-6

Carter, J. (2019). Philosophy of Mathematical Practice — Motivations, Themes and Prospects. *Philosophia Mathematica*, *27*(1), 1–32. https://doi.org/10.1093/philmat/nkz002

Casullo, A. (1988). Revisability, Reliabilism, and A Priori Knowledge. *Philosophy and Phenomenological Research*, *49*(2), 187–213. https://doi.org/10.2307/2107973

Cohen, D. I. A. (1991). The Superfluous Paradigm. In J. H. Johnson & M. Loomes (Eds.), *The Mathematical revolution inspired by computing* (pp. 323-329). Clarendon Press.

De Toffoli, S. (2021). Groundwork for a Fallibilist Account of Mathematics. *The Philosophical Quarterly*, *71*(4), pqaa076. https://doi.org/10.1093/pq/pqaa076

Gonthier, G. (2008). Formal proof–the four-color theorem. *Notices of the AMS*, *55*(11), 1382–1393. https://www.ams.org/notices/200811/tx081101382p.pdf

Hanna, G., & Larvor, B. (2020). As Thurston says? On using quotations from famous mathematicians to make points about philosophy and education. *ZDM*, *52*(6), 1137–1147. https://doi.org/10.1007/s11858-020-01154-w

Heesch, H. (1969). *Untersuchungen zum Vierfarbenproblem* (Vol. 810/810a/810b). Bibliographisches Institut.

Kitcher, P. (1984). *The nature of mathematical knowledge*. Oxford University Press. https://doi.org/10.1093/0195035410.001.0001

Kolata, G. B. (1976). The four-color conjecture: a computer-aided proof. *Science*, *193*(4253), 564–565. https://doi.org/10.1126/science.193.4253.564

Lean, O. M., Rivelli, L., & Pence, C. H. (2023). Digital Literature Analysis for Empirical Philosophy of Science. *The British Journal for the Philosophy of Science*, *74*(4), 875–898. https://doi.org/10.1086/715049

Löwe, B., & Kerkhove, B. V. (2019). Methodological Triangulation in Empirical Philosophy (of Mathematics). In A. Aberdein & M. Inglis (Eds.), *Advances in Experimental Philosophy of Logic and Mathematics* (pp. 15-38). Bloomsbury. https://doi.org/10.5040/9781350039049.0005

MacKenzie, D. (1995). The automation of proof: a historical and sociological exploration. *IEEE Annals of the History of Computing*, *17*(3), 7–29. https://doi.org/10.1109/85.397057

MacKenzie, D. (1999). Slaying the Kraken. *Social Studies of Science*, *29*(1), 7–60. https://doi.org/10.1177/030631299029001002

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. https://doi.org/10.1017/cbo9780511809071

McEvoy, M. (2007). The Epistemological Status of Computer-Assisted Proofs. *Philosophia Mathematica*, *16*(3), 374–387. https://doi.org/10.1093/philmat/nkn014

McEvoy, M. (2022). Revisiting 'The New 4CT Problem'. In *Handbook of the History and Philosophy of Mathematical Practice* (pp. 1-21). Springer. https://doi.org/10.1007/978-3-030-19071-2_38-1

McKinney, M. (2011). arXiv.org. *Reference Reviews*, *25*(7), 35–36. https://doi.org/10.1108/09504121111168622

Mejía-Ramos, J. P., Alcock, L., Lew, K., Rago, P., Sangwin, C., & Inglis, M. (2019). Using Corpus Linguistics to Investigate Mathematical Explanation. In *Methodological Advances in Experimental Philosophy* (pp. 239-264). Bloomsbury Academic. https://doi.org/10.5040/9781350069022.ch-009

R, C. T. (2024). R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Resnik, M. D. (1999). *Mathematics as a Science of Patterns*. Oxford University Press. https://doi.org/10.1093/0198250142.001.0001

Robertson, N., Sanders, D., Seymour, P., & Thomas, R. (1997). The Four-Colour Theorem. *Journal of Combinatorial Theory, Series B*, *70*(1), 2–44. https://doi.org/10.1.1.158.9438

Rota, G.-C. (1997). The Phenomenology of Mathematical Proof. *Synthese*, *111*(2), 183–196. https://doi.org/10.1023/A:1004974521326

Stewart, I. (1981). Concepts of Modern Mathematics. Penguine Books.

Swart, E. R. (1980). The Philosophical Implications of the Four-Color Problem. *The American Mathematical Monthly*, *87*(9), 697–707. https://doi.org/10.2307/2321855

Tanswell, F. S., & Inglis, M. (2024). The Language of Proofs: A Philosophical Corpus Linguistics Study of Instructions and Imperatives in Mathematical Texts. In *Handbook of the History and Philosophy of Mathematical Practice* (pp. 2925-2952). Springer International Publishing. https://doi.org/10.1007/978-3-031-40846-5_50

Thurston, W. P. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society*, *30*(2), 161–178. https://doi.org/10.1090/s0273-0979-1994-00502-6

Tymoczko, T. (1979). The Four-Color Problem and Its Philosophical Significance. *The Journal of Philosophy*, *76*(2), 57–83. https://doi.org/10.2307/2025976

Wilson, R. J. (2014). *Four colors suffice. How the map problem was solved* (Revised color edition. ed.). Princeton University Press.

Appendix

year	total	matches	%
1986	1	0	0%
1988	1	0	0%
1989	6	0	0%
1990	26	0	0%
1991	353	0	0%
1992	3.190	0	0%
1993	6.729	0	0%
1994	10.078	0	0%
1995	13.006	1	0,01%
1996	15.872	1	0,01%
1997	19.610	1	0,01%
1998	24.170	1	0%
1999	27.700	0	0%
2000	30.670	8	0,03%
2001	33.140	8	0,02%
2002	36.107	4	0,01%
2003	39.392	11	0,03%
2004	43.714	10	0,02%
2005	46.881	14	0,03%
2006	50.314	27	0,05%
2007	55.752	21	0,04%
2008	58.810	26	0,04%
2009	64.071	34	0,05%
2010	70.288	34	0,05%
2011	76.602	62	0,08%
2012	84.374	61	0,07%
2013	92.875	84	0,09%
2014	97.590	99	0,1%
2015	105.130	102	0,1%
2016	113.440	120	0,11%
2017	123.781	150	0,12%
2018	140.377	172	0,12%
2019	155.917	191	0,12%
2020	178.275	235	0,13%
2021	181.599	272	0,15%
2022	185.987	280	0,15%
2023	209.223	275	0,13%
2024	243.662	348	0,14%
Σ	**2.638.713**	**2.652**	**0,1%**

year	math total	math matches	cs total	cs matches	eess total	eess matches	other total	other matches
1986	0	0	0	0	0	0	2	0
1988	0	0	0	0	0	0	1	0
1989	6	0	0	0	0	0	0	0
1990	24	0	2	0	0	0	0	0
1991	61	0	3	0	0	0	308	0
1992	358	0	1	0	0	0	3.356	0
1993	603	0	7	0	0	0	7.837	0
1994	948	0	251	0	0	0	11.926	0
1995	1.198	1	255	0	0	0	15.989	0
1996	1.419	1	237	0	0	0	19.436	0
1997	1.849	0	192	0	0	0	23.468	2
1998	2.697	0	334	1	0	0	27.051	0
1999	3.348	0	325	0	0	0	30.431	0
2000	4.096	1	511	7	0	0	31.992	0
2001	4.354	1	613	5	0	0	34.551	2
2002	5.637	2	702	1	0	0	36.692	1
2003	6.648	2	865	0	0	0	39.430	9
2004	8.352	5	1.022	1	0	0	42.359	4
2005	10.011	8	1.385	4	0	0	44.337	6
2006	11.996	3	1.898	12	0	0	46.114	13
2007	14.242	6	2.842	13	0	0	49.148	4
2008	15.518	6	3.645	20	0	0	50.851	2
2009	17.585	9	4.875	20	1	0	54.217	6
2010	21.172	12	7.585	25	3	0	61.589	4
2011	24.162	19	9.125	47	5	0	67.199	8
2012	27.245	14	12.335	52	6	0	72.298	6
2013	30.223	15	14.940	72	7	0	75.448	9
2014	32.103	27	16.312	83	4	0	77.397	11
2015	34.737	32	18.834	79	23	0	81.633	5
2016	36.395	38	23.712	83	36	0	85.007	8
2017	38.483	40	30.813	119	750	0	87.944	9
2018	40.034	31	41.950	144	3.677	2	98.102	15
2019	42.983	44	55.524	163	9.842	24	109.903	25
2020	46.100	48	71.423	204	15.706	28	120.318	25
2021	45.135	85	77.517	223	14.381	32	111.901	34
2022	45.410	74	82.108	235	15.837	32	111.266	20
2023	47.924	74	100.015	240	17.256	25	117.659	19
2024	51.949	97	125.425	292	19.511	27	127.055	24
Σ	**675.005**	**695**	**707.583**	**2.145**	**97.045**	**170**	**1.974.215**	**271**

keyword	**matches**	**%**
acl2	97	3,66%
agda	188	7,09%
computer assisted	232	8,75%
computer-assisted	734	27,68%
coq	607	22,89%
hol	93	3,51%
idris	17	0,64%
isabelle	380	14,33%
lean	233	8,79%
metamath	21	0,79%
mizar	50	1,89%
Σ	**2.652**	**100%**