Mathematical proofs in practice: Revisiting the reliability of published mathematical proofs

Abstract

Mathematics seems to have a special status when compared to other areas of human knowledge. This special status is linked with the role of proof. Mathematicians often believe that this type of argumentation leaves no room for errors and unclarity. Philosophers of mathematics have differentiated between absolutist and fallibilist views on mathematical knowledge, and argued that these views are related to whether one looks at mathematics-in-the-making or finished mathematics. In this paper we take a closer look at mathematical practice, more precisely at the publication process in mathematics. We argue that the apparent view that mathematical literature, given the special status of mathematics, is highly reliable is too naive. We will discuss several problems in the publication process that threaten this view, and give several suggestions on how this could be countered.

Keywords: Mathematical proof, Reliability, Publication process, Absolutism, Fallibilism

1 Introduction

In recent decades a growing number of philosophers of mathematics have relocated their attention from the logical foundations of mathematics to the practice of mathematics[1]. In this paper we proceed along these lines by critically reevaluating the reliability of the publishing process in mathematics.

In section 2 we will discuss how mathematics is often characterized as epistemically unique, and relate this view with absolutist and fallibilist approaches to mathematics. Our goal is to show that the argument that the view that mathematics is epistemically unique does not necessarily lead to a high level of reliability of publications in mathematical literature. We will argue in section 3 why this is the case. Section 4 discusses possible strategies to counter the problematic situation that we outlined in section 3. Section 5, finally, entails some concluding remarks.

[1] See, for instance, the work of Mancosu (2008) and the recently founded Association for the Philosophy of Mathematical Practice or APMP (http://institucional.us.es/apmp/index.htm). For more insights on the historical development of the study of mathematical practice and its relationship with traditional philosophy of mathematics, see Van Bendegem (2014).

2 Mathematical knowledge

2.1 Epistemic uniqueness

Mathematical knowledge is often characterized as very special and unique, since it seems that mathematical knowledge has a level of certainty, exactness and objectivity that is unachieved in any other form of knowledge, including scientific knowledge. Mathematicians defend such a view with confidence. Kline argues as follows: “Whenever someone wants an example of certitude and exactness of reasoning, he appeals to mathematics.” (Kline, 1980, p. 4) Krantz also describes the idea that mathematics “is equipped with a sort of certainty that other sciences do not possess. [. . .] We have endowed the system with a reliability and reproducibility and portability that no other science can hope for.” (Krantz, 2011, p. 22)

Philosophers have wondered about the special epistemic status of mathematics as well. Heintz (2000) refers to two typical characteristics of mathematics, i.e. coherence of mathematical concepts and theories and the consensus among mathematicians. First, mathematics is strongly connected and coherent. What mathematicians and scientists have in common is that they work relatively isolated and within small specialized subfields. In contrast with science, where this practice leads to partly contradictory theories, mathematics remains highly coherent. Second, Heintz refers to a high consensus among mathematicians, or, as she indicates, mathematics does not allow flexibility on the part of interpretation. More specifically, mathematics leaves no room for controversy in regard to its conclusions. Anyone accepting the rules of the mathematical method is thus supposed to arrive at the same result.

2.2 Absolutism and fallibilism in mathematics

Several philosophers have nonetheless pointed out that the above view of mathematics is mistaken, since it it is not a correct or complete presentation of the mathematical practice or mathematical knowledge. Ernest (1998) describes two opposing views in philosophy of mathematics, that is the absolutist and the fallibilist view about mathematical knowledge. Absolutism sees mathematical knowledge as a body of knowledge consisting of certain and unchallengeable truths. Sympathisers of the absolutist view argue that the use of formal language, axioms and strict rules of inference in mathematics leads to unquestionable mathematical knowledge. Furthermore, they argue that these tools suffice to establish all mathematical truths.

The other view Ernest describes, and argues for, is fallibilism. This is a more recent position in the philosophy of mathematics, and embraces the practice of mathematics and its human side. Mathematical knowledge is here described as fallible and corrigible, both in terms of its proofs and its concepts. Lakatos (1978) is often considered among the first and most important authors that presented such a fallibilistic view. In *Proofs and Refutations*, he presents a lively classroom discussion of Euler’s formula in order to explore ideas about the concept of proof and the nature of mathematics. By showing how the formula was investigated, decomposed, built upon and improved, Lakatos presented a mathematical activity that was an alternative to the idea that mathematics was merely presenting an absolute proof. Through his analogy, Lakatos demonstrated mathematics as an activity of trial and error. Mathematical concepts, forms and standards of proof are not permanently fixed, but can rather have a history of continuous modification.

One way of approaching these discussions is by differentiating between writing mathematics and doing mathematics. Hersh (1997) has called this respectively the 'front' and the 'back' of mathematics. The analogy is made by referring to a gourmet restaurant, where in the front the public are served finished and perfect dishes. The back, which is the kitchen, is restricted to insiders and shows a much more messy (but often pratical) activity. Hersh argues that in mathematics an analogous distinction between the front, how mathematics is presented neatly in lectures and textbooks, and the back, how mathematicians 'make' new knowledge and how this is connected with a messy human striving. The front still leaves the impression of an absolutist view of mathematics while the back shows a more fallibilist picture.

2.3 Mathematical publications

The dichotomy between absolutism and fallibilism has mainly led to extensive discussions in the philosophy of mathematical education, on how mathematics should be presented in the classroom. Within the community of working mathematicians, however, such discussions are quite rare [2]. In this paper, we examine a particular part of the mathematical practice, namely the publication process. It is not surprising that, in accordance with the work of Hersh, there are differences between the activities that are part of creating a mathematical idea and the properties of the final paper. We are interested in the latter, and whether one could still uphold the idea that knowledge presented in mathematical literature reaches an uniquely high level of reliability or the knowledge presented in papers is in fact part of the fallible striving towards true knowledge.

Mathematicians do not merely present the claims or conclusions of their research in published articles. They use the method of proof, in order to show how the truth of these claims is mathematically established. A crucial aspect of proof seems to be that, if two or more mathematicians share the same mathematical assumptions and reasoning methods, one can check a proof of his colleague completely in order to verify the correctness of the mathematical result.

Obviously, working mathematicians are humans and are not immune to error. This is the case even if mathematics offers this method that, in principle, leaves no room for errors. Mathematical research is complex. There is always a possibility that mathematicians make a mistake that they oversee themselves. Consequently, if a mathematician presents a mathematical result, it is possible his or her research is infected by mistakes.

[2] The gap between concepts in (philosophy of) mathematical education and mathematical practice is discussed by Sfard (1998)

The most common and important medium mathematicians use in order to present these results, in agreement with other sciences, is the publication in academic journals. This sort of communication has two specific characteristics. First, these journals are peer reviewed. As a result, every paper is reviewed by at least one mathematician before it is accepted for publication. Secondly, after the publication anyone with access to the journal can read the article.

Starting from these assumptions one would be able to argue that the mathematical literature is highly reliable. If an article contains an error, due to a mistake of the author, this error can be detected by the reviewers of the journal. Even if these reviewers do not detect the error, other mathematicians will be able to detect it. This is the case even when a mistake eludes detection for a long period, which is a characteristic that makes mathematics unique, as discussed by Azzouni: "[E]ven if many results are built on that mistake, this wont provide enough social inertia once the error is unearthed to resist changing the practice back to what it was originally: in mathematics, even after lots of time, the subsequent mathematics built on the falsehood is repudiated." (Azzouni, 2007, p.9)

In the following section we argue that in contemporary practice, this argument is too naive. Although we believe that these assumptions have a certain truth in them, the reality is more complex and problematic.We will argue that the reliability of mathematical literature should be handled with greater caution than is done in contemporary practice. The result is that if we want to deal with this problem, as we will do in the final section by suggesting several strategies, that mathematical literature should be more located in the back and treated in terms of the fallibilist.

3 Mathematical Practice

3.1 Surveyability of mathematical proofs

A first important observation is that mathematicians, in contrast with other sciences, are able to prove their results. Once such a proof is available, another mathematician can in principle check the correctness of the proof without any doubt, by making sure that every part of the proof follows legitimately from the previous ones. The question that evidently arises is whether every proof can be checked this rigorously. These questions are linked with the topic of surveyability, that has been severely discussed in mathematics and philosophy [3].

Contemporary discussions on the surveyability of mathematical proofs start with the seminal paper on the four-color theorem by Tymoczko (1979). The problem is that the proof, provided by Appel and Haken, is obtained with the use of computer programs. The data used is too immense in order to be checked by a mathematician. Tymockzo argued that if we accept the four-color theorem based on a computer-assisted proof we have to change the concept

[3] For a historical overview, see (Bassler, 2006)

of proof. The standard view of proof was indeed a "construction that can be looked over, reviewed, verified by a rational agent. We often say that a proof must be perspicuous, or capable of being checked by hand. It is an exhibition, a derivation of the conclusion, and it needs nothing outside itself to be convincing. The mathematician surveys the proof in its entirety and thereby comes to know the conclusion."(Tymoczko, 1979, p. 59) Tymoczko thus showed, that the fact that most of the mathematical community did accept the proof as a genuine mathematical justification for the four-color theorem, meant a crucial change in traditional conceptions of proof and mathematical knowledge.

Contemporary mathematics is increasingly reliant on the interaction with computer systems, and not only for reasoning or calculation steps in proofs. There are also more and more programs that verify mathematical results.[4] In order to use these tools, the proof must be written in a completely formal language, which is a precise artificial language that only admits certain well-defined operations. Naturally, the proof-checking software could contain bugs and again, mathematicians are not always able to check everything themselves because the processed data is too immense. There is however another problem, namely that the proofs mathematicians deal with are almost never spelled out in a complete formal language: "Human mathematics consists in fact in talking about formal proofs, and not actually performing them. One argues quite convincingly that certain formal texts exist, and it would in fact not be impossible to write them down. But it is not done: it would be hard work and useless because the human brain is not good at checking that a formal text is error-free. Human mathematics is a sort of dance around an unwritten formal text, which if written would be unreadable." (Ruelle, 2000, p. 254)

Hence, for a set of (non-traditional) proofs the certainty of the established theorem is not obtained by referring to the surveyability of its proof. But at least most mathematicians are willing to discuss what place they give these theorems in the body of mathematical knowledge. More interesting and more problematic are the traditional proofs. Perhaps a move towards proof-checking programs is needed, but we are in this paper interested in the proofs that are in fact published. And these proofs are not presented in a way that the surveyability and verifiability is trivial[5].

Fallis (2003), for instance, introduces the notion of gaps in mathematical proofs [6]. A gap is any point where the written proof does not follow from the previous lines in the proof by merely applying formally valid rules of inference. Fallis argues that most actual proofs that are presented by the mathematical community contain gaps. He proposes the following categorization of proof gaps: A mathematician leaves an *inferential gap* whenever the mathematician

[4]For further discussions of this topic, see (Harrison, 2008).

[5]It could be that certain informal mathematical methods, such as mathematical visualizations, are tools to verify mathematical results as well. Giaquinto (2007), for example, argues that mathematical visualisations can lead to legitimate mathematical knowledge. Further investigations in these topics are more than welcome.

[6]Fallis introduces this notion in order to argue that probabilistic proofs in mathematics should be given more credence.

uses a particular sequence of propositions, which he/she has in mind as being a proof, that is in fact not a proof. This sort of gap is problematic, as it undercuts the proof. A mathematician leaves an *enthymematic gap* if he/she does not explicitly state the particular sequence of propositions in mind. One has supposedly checked all the details, but these details are left out for the sake of style or length of the article. The third category are *untraversed gaps*, where the mathematician has not tried to verify directly that each proposition in the sequence of propositions that he/she has in mind follows from a subset of previous propositions in the sequence by application of a mathematical inference. Fallis notes that in some cases it is "considered acceptable for a mathematician to leave an untraversed gap." (Fallis, 2003, p. 59) Next to this taxonomy of gaps, Fallis speaks of a universally untraversed gap if none of the members of the mathematical community has bridged an untraversed gap. He claims that such gaps are not unusual in mathematical practice and that proofs containing such gaps are still often accepted/justified by the mathematical community [7].

A similar point, using another terminology, is also made by Davis (1972), who discusses the practice of splicing proofs: "In the course of a proof, one cites Euler's Theorem, say, by way of authority. The onus is now on the reader to verify that all the conditions (in their most modern formulation) which are necessary for the applicability of the theorem are, in fact, present." (Davis, 1972, p. 259) Splicing allows mathematicians to leave out steps that are sufficiently worked out in other places of the literature. Davis argues that splicing is not the only phenomenon in mathematical practice. Mathematicians also skip steps. The practice of skipping is very similar to the taxonomy of gaps provided by Fallis. Davis states that mathematicians skip steps due to boredom, if he or she finds it unnecessary to complete every single step, or superiority, if he or she believes that everyone can follow the steps without spelling them out.

Mathematicians thus leave, in practice, gaps in the communication of mathematical proofs. On the one hand this makes the proofs more readable, on the other hand it is not always clear to what extent the proofs can still be checked rigourously. Coleman (2009), discussing the surveyability of long proofs in mathematics, concludes that a dichotomy between short and long proofs is unnecessary since mathematics always concerns long proofs, emphasizing the importance of the archive of mathematical literature in this matter: "The key to that, is to recognise that mathematics is a written practice which depends on the accumulation and deployment of an archive. Because of that, one does not need to have all of a proof in ones head because it is all on record. Understanding it means you have the main idea of it in your head, and knowing it requires both that, and the facility with the archive to get at the details if wanted. The fully written out proofs are there, in a sense - the distributed sense in which the mathematically educated agent can access them, and so the theorems of mathematics are indeed justified by proofs, though the true justifications are

[7] Fallis admits that it is difficult to cite cases where it is undeniable that a mathematician has left a universally untraversed gap and that he/she is nevertheless justified in his/her belief in a a manner that is accepted by the mathematical community. Fallis discusses several cases that suggest that leaving universally untraversed gaps is at least sometimes accepted.

not the proofs labeled as such. Really, mathematics is long proofs." (Coleman, 2009, p. 42)

We believe that this is a good view of mathematical research, but we want to include two remarks. First, not all "fully written out proofs" are available, since published proofs are not only sliced, but certain steps are also skipped. And secondly, it shows that checking a mathematical proof comes with a high price. If one wants to fully check a mathematical result, one has to consult the archives of mathematical literature and fill in certain gaps in addition to reading the paper itself. The idea that mathematical proofs can be checked beyond doubt given common assumptions and methods between mathematicians, is thus not as simple as it appears. Moreover, as we will discuss in the next sections, proofs are not sufficiently checked at all.

3.2 Peer review

Once a paper containing mathematical results is written, it is submitted to an academic journal. Before it can be published, the article must be reviewed by referees. These referees, other professional mathematicians who are experts in the area covered in the article, decide whether the article can be published in the journal. One of the goals of the peer review system is quality control. It protects the body of literature from contamination by false or flawed arguments.

Does the refereeing process lead to a high reliability of mathematical literature? Although we have seen that this is also problematic, the use of (semi-)verifiable proofs could still lead to a positive answer. Nathanson, in an opinion piece for a mathematical journal is however quite pessimistic on this: How do we know that a proof is correct? By checking it, line by line. [. . .] If a theorem has a short complete proof, we can check it. But if the proof is deep, difficult, and already fills 100 journal pages, if no one has the time and energy to fill in the details, if a complete proof would be 100,000 pages long, then we rely on the judgment of the bosses in the field. [. . .] Many (I think most) papers in most refereed journals are not refereed. There is a presumptive referee who looks at the paper, reads the introduction and the statement of the results, glances at the proofs, and, if everything seems okay, recommends publication. Some referees check proofs line-by-line, but many do not. When I read a journal article, I often find mistakes. Whether I can fix them is irrelevant. The literature is unreliable." (Nathanson, 2008, p. 773)

Nathanson mentions several problems. First, mathematicians are not able to check all the proofs themselves. We will discuss this problem more severe in the next section. Second, referees do not check the correctness of the proofs. Hildebrand (2009) also states that reviewers can only be reasonably confident about the correctness of the article. It remains however vague what this amounts to and whether this happens in practice. And third, Nathanson (2008) was able to find mistakes in the literature. Nathanson does not stand alone with his claim that the literature contains error. Davis (1972) states that half of the published proofs contain errors [8].

Geist, Löwe and Van Kerkhove (2010) have done preliminary research on the status of the refereeing process and the link with the reliability of mathematical literature. They remarkably found that it is not even universally expected that referees check the complete correctness of the paper: "Mathematicians disagree about the amount of detail checking that has to be done by the referees. While some (few) mathematicians think that checking the correctness of the proofs is the main task of the referee, others disagree with this and consider mathematical correctness the problem of the author rather than that of the referee." (Geist et al., 2010, p. 161) This is confirmed by a questionnaire the authors sent off to several mathematical journals. One of the questions was whether the referee should check all the proofs in detail, some proofs in detail or no proofs in detail. Of the 27 editors that were addressed, 11 answered: the first option was chosen five times, the second six times, and the third option was not selected. Remarkably, one of the editors who chose the first option added: "but to be reasonable, I am happy when I find a referee doing (b)"(Geist et al., 2010, pp. 163-164)), which is in fact the second option.

3.3 Mathematical community

Mathematicians, as mentioned above, base their work on the archive of other mathematical results. Consequently, it can still be the case that mathematical publications are sufficiently checked when they are used for new research. However, this view is still quite naive, as already became clear in the quote from Nathanson in the previous section. Mathematicians do not have the time or expertise to completely check every proof they read or use. Rather, Geist, Löwe and Van Kerkhove (2010) state there is a tendency of mathematicians to take proofs for granted in using them for their own research.

The fact that knowledge depends on the knowledge reported by others is discussed in the topic of testimony[9]. It is evident that a great number of our beliefs come from what others tell us. The epistemology of testimony takes up the challenge to investigate when and why these beliefs are justified.

Hardwig (1985)(1991) acknowledges trust as an ultimate foundation for a serious part of our knowledge[10]. He argues that this trust is eligible since "if I were to pursue epistemic autonomy across the board, I would succeed in holding relatively uninformed, unreliable, crude, untested and therefore irrational beliefs." (Hardwig, 1985, p. 340) Mathematical progress would indeed be hard

[8] There is naturally a difference between errors that undermine the established theorems, and errors that are present in the paper but don't render the results false. There is no reason to believe that the majority of mathematical results in the literature is false because there are quite a lot of errors. This could be linked with Thurston's (1994) claim that mathematicians are manifest better in detecting flaws or weaknesses that make a proof false than validating its complete formal correctness.

[9] For a detailed overview of both historical and contemporary positions in this debate, see (Origgi, 2004).

[10] For insights on how its counterparts, being distrust and mistrust, influence knowledge gaining, see (Primiero and Kosolosky, 2014)

without relying on testimony of others. It would, for instance, require a lot of time and expertise for a mathematician to check the reliability of each accepted proposition in a proof. Moreover, the proofs that the mathematician would be able to check, start themselves from already accepted propositions. Despite the fact that testimony seems to be an essential part of mathematical research and progress, we should still ask ourselves when relying on testimony is in fact justified. Or, in other words, when we are justified taking the work and word of our peers for granted in a rather 'blind' manner, as we miss out on time and energy to justify their claims for ourselves.

The mathematician Auslander refers to the importance of experts in this matter. Mathematicians take certain results for granted, without verifying the proofs themselves: "We accept that a purported result is correct when we hear that it has been proved by a mathematician we trust and "validated" by experts in the authors mathematical specialty. This is the case even if we havent read the proof, or more frequently when we dont have the background to follow the proof."(Auslander, 2008, p. 64)

Furthermore, many mathematicians adopt the view that if a mathematical result is published in a journal, it must be trustworthy. Weber (2008) (2011) interviewed several mathematics professors at a regional university in the United States, specialized in a variety of mathematical disciplines. On the topic of validating the correctness of the proofs, one of the interviewees said: "To be honest, when I read papers, I don't read the proofs. In the journal papers, and the papers that I read in my research, maybe that's bad or maybe that's not, if I'm convinced that the result is true, I don't necessarily need to read it, I can just believe it." (Weber, 2008, p. 449) Another professor said the following: "I do not try and determine if a proof is correct. If it is in a journal, I assume it is. I'm much more interested in the ideas of a proof." (Weber and Mejia-Ramos, 2011, p. 334) It must be noted that it is not the case that all participants indicate that they do not check proofs in the literature. In fact, some indicate that they do pay attention to the correctness of proofs when they read papers. Nevertheless, these quotations show that Nathanson and Auslander are not mistaken when indicating that checking the correctness of a proof is not the primary concern of mathematicians. And we already discussed the problem of depending on peer review for the correctness of these proofs.

Another observation connected with the problem of testimony is the fact that, which may seem surprising for an outstander[11], some mathematical results in fact lack a written proof. This is, among others, described by the mathematician Thurston: "Within any field, there are certain theorems and certain techniques that are generally known and generally accepted. When you write a paper, you refer to these without proof. You look at other papers in the field, and you see what facts they quote without proof, and what they cite in their bibliography. [...] Many of the things that are generally known are things for which there may be no known written source. As long as people in

[11] Mathematicians themselves even have a name for proofs that may circulate without having been appeared in print, namely mathematical folklore.

the field are comfortable that the idea works, it doesn't need to have a formal written source." (Thurston, 1994, p. 168)

The attitude of taking other literature for granted does not only stand in the way of detecting and correcting errors in the literature. In reality, it leads to new results that are based on flawed literature. Consequently, it becomes increasingly complex to correct mistakes in the literature, since new results are based on other results that are potentially flawed and potentially insufficiently checked by either reviewers or authors of the new article.

3.4 Mathematical journals

A final issue is that authors are in fact not always interested in validating other results, and more interested in finding new results. In science, authors may be inclined to publish statistically significant results more eagerly than non-significant ones, and negative findings are more difficult to prove with a high probability than positive ones (Resnik, 2007). Moreover, results that are more provocative may get published more easily, as a study on precognition demonstrates: Bem (2011) published an article in which he reported evidence of peoples ability to see the future. Ritchie, French and Wiseman (2012), who replicated the study, found, however, no evidence of precognition. Their paper was however rejected by the same journal as Bem's research was published in. The journal mentioned that their policy was not to publish replications of previous studies. The replication study was also turned down by Science, only now the reason was that the negative results should be published in the same journal as the original study, which the authors previously attempted to do. Psychological Science and Science Brevia is yet another journal that rejected the study without conducting proper peer review (Goldacre, 2011)(Sutton, 2011). Our example shows how difficult it is to get negative studies published. Science is thought to be self-correcting partly because there is the possibility of replicating previous studies. However, while researchers are pressured to publish frequently as to promote their career, the editors policy of not publishing replication studies does not encourage them to take part in the process of correcting each others work. This, in turn, can hinder the detection of biases and mistakes in earlier studies (Jukola, Forthcoming).

Grcar (2013) has discussed the fact that, compared to other areas of investigation, mathematical literature has significantly lower correction rates. However, considering the previous sections, it is hard to maintain that mathematical literature has in fact lower error rates than these other areas of investigation. One possible explanation is that mathematicians are in fact absolutists that believe that papers that undergo peer review are free from error. This point has been discussed earlier. Another explanation that Grcar points out is a lack of editorial guidance: "The lack of editorial policies inviting correction goes hand in hand with low rates in mathematical journals. [...] Some journals privately advise authors to post corrections on their personal websites, which is not consistent with the concept of an archival literature." (Grcar, 2013, p. 422) The idea that mathematical research is based on the archive of literature indeed demands that errors in the archive are not only detected, but that these errors are also indicated and/or corrected in the archives themselves. This is not the case, because similarly to the discussed case in psychology, journals are not eager to publish such results.

An example of this problem is an online article of Capelas de Oliveira and Rodrigues Jr (2006). They provide a comment on a published paper on computations in multi-dimensional mathematics. The authors of the comment provide a severe criticism and several counterexamples to what they take to be unbelievable results of the published paper. However, this valid comment has not been published, while the concerning paper is not retracted and still available in the archival literature. A second example is a letter from Hill(2010), who also found a sufficient amount of counterexamples to a published mathematical paper. In a letter to the journal he complains about the reluctance of publishing his remarks fully, and the problematic effects of the fact that the paper is not retracted or corrected sufficiently: "The Editorial solution was a "correction" that not only omitted many of the flaws I had found, but also introduced a new mathematical error into the permanent record in spite of an additional counterexample I had provided them to the new claim. The same errors that appeared in the 2006 Notices article continue to be propagated in the literature, including published research articles and a new book. More than two years later, I am still trying to publish the counterexamples, and will continue to do so." (Hill, 2010, p. 7)

4 Discussion

It is not our goal to argue that relying on testimony, or, in other words, 'blindly' trusting one's peers, in mathematical practice should be prohibited, since it would make mathematical research and progress impossible. However, the idea of working mathematicians that published articles are free from error creates an absurdly high expectation of peer review to catch all errors. We will argue in this section for more transparency and a fair credit allocation, in order to encourage mathematicians to detect, acknowledge and discuss mistakes in the literature.

4.1 Plea for transparency

Increasing the level of transparency breaks the taboo that mathematical knowledge in the literature is as infallible as an absolutist view suggests.

A first way to achieve more transparency is using the possibilities and potential of the internet fully. Several sciences already began to adapt to the 'new' medium by building and using online databases that were designed in order to share data among a community of scientists. An example is the Gene Expression Omnibus[12]. As an answer to the demand for a public repository for data generated from high-throughput microarray experiments, this database allows scientists to submit, store, retrieve and consult many types of data sets.

[12] Which can be found online on: http://www.ncbi.nlm.nih.gov/geo/

There are also examples from the field of mathematics. A first case is the Metamath Proof Explorer[13]. Inspired by Whitehead and Russel's project Principia Mathematica, the aim of this database is to collect completely worked out proofs, starting from the very basic foundations of mathematics. As a result, each proof can be brought back to the axioms of logic and set theory. Many mathematicians would argue that the phenomena of slicing, splicing and gaps discussed earlier in this paper are in fact tools in order to avoid such proofs that demand lots of patience for even quite simple or basic proofs. But one can hardly say that such projects are not welcome in mathematical practice, where the complete correctness of mathematical literature is unattained. A second case is the Polymath and Tricki Project, two projects initiated in 2009 by mathematician and Fields Medallist winner Gowers. In January 2009 Gowers posted a unsolved mathematical problem on his blog and invited collaboration from anyone who was interested. About 40 people accepted the challenge and together they solved the problem within several weeks. This pace was highly uncommon for such a problematic case. The resulting proofs were eventually published under the pseudonym Polymath. Next to the Polymath project, Gowers launched Tricki[14], which is a Wikipedia-style database of articles about mathematical techniques that are useful in mathematical problem solving.

We argue for more and specific databases that increase the transparency on how a fully written out proof could be completed. Given the length of some proofs, and the archive of proofs of assumptions of the new proof, it is indeed impossible and unnecessary to include every step in the literature. If such an omission is the case, for any reason whatsoever, an online database could serve as an archive for the steps that the author of the article has in fact completed. An online database could serve as a substitute. If there are steps that (s)he does not complete because it is too trivial, the author should mention this. If (s)he leaves out steps because they are in the literature (either in work of the author self, or in other work), (s)he should include references to these proofs. And finally, if (s)he leaves out steps because (s)he has done them, but are left out in function of the length, comprehensibility or style of the article, (s)he should put them in the database. We hope the benefits of such a database are clear. If a mathematician has in fact gone through some steps, in a formal or informal way, it is very little effort to put them in the database. Consequently, it is more transparent, in case of doubt, to check steps or fill in gaps of the proof. The success of Gowers' Polymath shows a goodwill among mathematicians to tackle mathematical problems collaboratively. As we have argued throughout the paper, the correctness of publications should be seen as a challenge as well.

Not only authors should be more transparent, but peer reviewers as well. Again, we will not argue that the responsibility of the correctness of the proofs lays fully with the peer reviewers. Such a view is unattainable, since we agree that reviewers can in fact only be reasonably certain about the correctness of the proof. However, we do defend the obligation of the peer reviewers to be

[13] Which can be found online on: http://us.metamath.org/mpegif/mmset.html

[14] Which can be found online on: www.tricki.org

transparent about their contribution to checking the correctness of proofs. If a peer reviewer checks certain steps, while (s)he is not able to fully check another part of the proof, for example due to his/her area of expertise, this can and should be public information. Such information is useful and in fact crucial for other mathematicians, whether they are other reviewers, editors or readers.

4.2 Credit allocation

Looking into how scientific research is socially organized has often led to the characterization of a researcher as a credit-maximizing individual. Research is driven rather by the search for credits than a search for truth. Researchers can for example gain credits by publishing results in respected journals. Consequently, such credits lead to new funding or job opportunities, which can be again formulated in terms of credit. The notion of credit has been used in philosophy of science, for example in order to explain how scientists choose certain research programs (Kitcher, 1993). Naturally, mathematicians are also interested in their academic positions and career, and mathematical practice can also be considered as a credit-driven practice.

The idea of credit-driven mathematics allows us to formulate several strategies that would lead to a situation where mathematicians can earn credit for checking proofs. This is now not the case, as refereeing is done anonymously. If referees are no longer anonymous, it would be possible to grant them credits for checking the correctness of proofs. It is hard to see why a referee that checks the correctness of proofs should be anonymous. Judgment on originality and publication worthiness of a paper can be another case, but perhaps these tasks should then be separated in the refereeing process.

It is also necessary to endorse an editor mentality shift concerning corrections of published papers. If such papers become more acceptable in the literature, mathematicians will be more inclined to take the time to check or complete other proofs. Now, as we have seen, such endeavors end up in internet databases or letters to journals, which leave the authors with no credits. Another possibility is that we devise a mathematical journal that solely focuses on contributions concerning the correctness of (already published) proofs. It is easy to see that for an innovative or highly difficult result in mathematical research the mathematician in question deserves credit. But the previous sections have shown that contributing results about a flaw or weakness of a proof, filling in gaps of a proof or asserting the correctness of a long or difficult proof are also important.

Mathematicians should earn credit for these endeavors as well. The possibility of gaining academic credit for checking proofs, encourages mathematicians to invest time in such tasks. In fact, it seems that in the current situations credits are what pushes them away from performing such tasks where no credit can be gained. Furthermore, more transparency, as discussed earlier, should also encourage mathematicians to take up these tasks. This is a scenario that could easily lead to an increased degree of reliability concerning the proofs and results established in mathematical literature .

Finally, our approach creates room to implicate the importance of a new, what Shirky (2008) calls a "publish-then-filter", approach to mathematical practice. Publish-then-filter, in our case, entails that mathematical journals publish proofs[15] knowing that they might turn out to be false later on[16], as opposed to postponing publication as long as one is not entirely sure about the reliability of its content. This idea could function in mathematical research as it already does in other scientific research. In academia ones claims get rejected all the time turning rejection into an intrinsic and accepted part of scientific research. Papers are but rarely withdrawn from previous publication, even if they contain evident mistakes. The reasoning behind it is that these mistakes are thought to be eventually filtered out through subsequent publications or meta-analyses or, as we suggested, through journals focusing solely on correcting flaws in previous papers. Not only is this approach closer to how scientific practice works, it would also take a considerable load off of reviewers worries. They should check the papers to a reasonable level, being transparent about what they have checked completely and what not. After publication, other mathematicians should than challenge the results further. This, however, can only work if mathematicians and the broader audience adjust their attitude towards mathematics (and knowledge-gaining in general), i.e. that results can be flawed and that this is part of the game.

What our scientific endeavors will bring us, we cannot predict. What we can predict, however, is that sticking to a rigid view on the epistemic uniqueness of mathematics, in terms of it possessing old-fashioned certainty, will certainly not be conducive. In this paper we made an attempt to adjust this view of mathematics and proposed strategies to tackle the problems that surround it. The alternative we propose is much more in line with fallibilism of mathematics, where processes from the back-checking, completing, changing, correcting,...-proofs are much more highlighted in the presentation of mathematical results.

5 Concluding remarks

In this paper we started with the differentiation between absolutist and fallibilist views on mathematical knowledge, and the argument that the view that mathematical knowledge is infallible is only defendable when one looks at the front of mathematics, as it is presented in lectures and books, and not when one looks at the back of mathematics, as how working mathematicians deal with problems in practice.

Starting from the assumptions that mathematicians are in fact humans and are not immune to error, one can still argue that the method of proof and communication through academic journals leads to a high level of reliability of mathematical publications. If an article contains an error, due to a mistake of

[15] Of course, we imagine this in accordance to our other suggestions made above.

[16] According to Shirky, its opposing strategy filter-then-publish is outdated, as "[it] rested on a scarcity of media that is a thing of the past. The expansion of social media means that the only working system is publish-then-filter." (Shirky, 2008, p. 98) For an impression how this idea would impact other disciplines, see (Kosolosky, Forthcoming).

the author, this error can be detected by the reviewers of the journal. Even if these reviewers do not detect the error, other mathematicians will be able to detect it. This is the case even when a mistake eludes detection for a long period.

However, we showed that, first of all, published proofs contain gaps and cannot simply be checked by humans. Secondly, at each level of the publication process, we argued that there is a general lack of discipline to emphasize the importance of checking and correcting published proofs. Given this state of the field, we came up with several ways in which mathematics, as a discipline, can and should be improved, so that it becomes more responsive to actual mathematical practice. These are (1) ways that increase the transparency among authors and reviewers, for instance through appealing to online databases, (2) ways that allow for appropriate credit allocation, both for authors (in the sense that venues are created for checking and correcting proofs) as for reviewers (in the sense that they get awarded for their work in checking proofs by making their names explicit) in a practice responsive system of publish-then-filter.

Working out these particular suggestions would require further research that goes beyond the scope of this paper. Nevertheless we managed to show that such ideas governed by transparency and credit allocation could help us to bring mathematical publications and mathematical practice closer together. At this point, we hope to have established some important questions for both philosophy of mathematics and the practice of mathematics.

References

Auslander, J. (2008). On the roles of proof in mathematics. In Gold, B. and Simons, R., editors, *Proof and other Dilemmas: mathematics and Philosophy*, pages 61–77. Mathematical Association of America, Washington DC.

Azzouni, J. (2007). How and why mathematics is unique as a social practice. In Van Kerkhove, B. and Van Bendegem, J., editors, *Perspectives on Mathematical Practice*, pages 3–23. Springer, Netherlands.

Bassler, B. O. (2006). The surveyability of mathematical proof: A historical perspective. *Synthese*, 148:99–133.

Bem, D. J. (2011). Feeling the future: Experimental evidende for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100:407–425.

Capelas de Oliveira, E. and Rodrigues Jr, W. (2006). A comment on "a comment on "on some contradictory computations in multi-dimensional mathematics". Online document.

Coleman, E. (2009). The surveyability of long proofs. *Foundations of Science*, 14:27–43.

Davis, P. J. (1972). Fidelity in mathematical dicsourse: Is one and one really two? *American Mathematical Monthly*, 79:252–263.

Ernest, P. (1998). *Social Constructivism as a Philosophy of Mathematics.* State University of New York Press, New York.

Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthese*, 134:45–69.

Geist, C., Löwe, B., and Van Kerkhove, B. (2010). Peer review and knowledge by testimony in mathematics. In Lwe, B. and Müller, T., editors, *PhiMSAMP: Philosophy of Mathematics: Sociological Aspects and Mathematical Practice*, pages 155–178. College Publications, London.

Giaquinto, M. (2007). *Visual Thinking in Mathematics.* Oxford University Press, Oxford.

Goldacre, B. (2011). I foresee that nobodoy will do anything about this problem. Online document.

Grcar, J. F. (2013). Errors and corrections in mathematics literature. *Notices of the American Mathematical Society*, 60:418–425.

Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82:335–349.

Hardwig, J. (1991). The role of trust in knowledge. *The Journal of Philosophy*, 88:693–708.

Harrison, J. (2008). Formal proof - theory and practice. *Notices of the American Mathematical Society*, 55:1395–1406.

Heintz, B. (2000). *Die Innenwelt der Mathematik, Zur Kultur und Praxis einer beweisenden Disziplin.* Springer, VIenna.

Hersh, R. (1997). *What is Mathematics, Really?* Oxford University Press, Oxford.

Hildebrand, A. (2009). Getting published in mathematics. Online document.

Hill, T. (2010). Hoisting the black flag. *Notices of the American Mathematical Society*, 57:7.

Jukola, S. (Forthcoming). Commercialization of research and the quest for the objectivity of science. In Froeyman, A., Kosolosky, L., and van Bouwel, J., editors, *Science versus Democracy: Social epistemology meets the philosophy of the humanities.* Foundations of Science.

Kitcher, P. (1993). *The Advancement of Science.* Oxford University Press, Oxford.

Kline, M. (1980). *Mathematics: The Loss of Certainty.* Oxford University Press, Oxford.

Kosolosky, L. (Forthcoming). Exploring how and why the ipcc blundered: Reassessing peer review in scientific practice. *Journal for General Philosophy of Science: Special Issue on Climate Change.*

Krantz, S. (2011). *The Proof is in the Pudding.* Springer, Netherlands.

Lakatos, I. (1978). *Proofs and Refutations.* Cambridge University Press, Cambridge.

Mancosu, P. (2008). *The Philosophy of Mathematical Practice.* Oxford University Press, Oxford.

Nathanson, M. (2008). Desperitley seeking mathematical truth. *Notices of the American Mathematical Society*, 55:773.

Origgi, G. (2004). Is trust an epistemic notion? *Episteme*, 1:67–72.

Primiero, G. and Kosolosky, L. (2014). The semantics of untrustworthiness. *Topoi*, DOI: 10.1007/s11245-013-9227-2.

Resnik, D. (2007). *Price of Truth.* Oxford University Press, Oxford.

Ritchie, S., Wiseman, R., and French, C. (2012). Failing the future: Three unsuccesful attempts to replicate bem's 'retroactive facilitation of recall' effect. *Plos One*, DOI: 10.1371/journal.pone.0033423.

Ruelle, D. (2000). Conversations on mathematics with a visitor from outer space. In *Mathematics: Frontiers and Perspectives.* American Mathematical Society.

Sfard, A. (1998). The many faces of mathematics: Do mathematicians and researchers in mathematics education speak about the same thing? *Mathematics Education as a Research Domain*, 4:491–511.

Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations.* Penguin Press, New York.

Sutton, J. (2011). Psi study highlights replication problems. Online document.

Thurston, W. P. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society*, 30:161–177.

Tymoczko, T. (1979). The four-color problem and its philphilosophy significance. *The Journal of Philosophy*, 76:57–83.

Van Bendegem, J. P. (2014). The impact of philosophy of mathematical practice on the philosophy of mathematics. In Soler, L., editor, *Rethinking Science after the Practice Turn.* Springer, Netherlands.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, 39:431–459.

Weber, K. and Mejia-Ramos, J. P. (2011). How and why mathematicians read proofs: An exploratory study. *Educational Studies in Mathematics*, 76:329–344.