Acceptable Gaps in Mathematical Proofs

Andersen, Line Edslev

Published in:
Synthese

DOI:
10.1007/s11229-018-1778-8

Publication date:
2020

Document Version:
Accepted author manuscript

Acceptable gaps in mathematical proofs

Line Edslev Andersen[1]

Abstract Mathematicians often intentionally leave gaps in their proofs. Based on interviews with mathematicians about their refereeing practices, this paper examines the character of intentional gaps in published proofs. We observe that mathematicians' refereeing practices limit the number of certain intentional gaps in published proofs. The results provide some new perspectives on the traditional philosophical questions of the nature of proof and of what grounds mathematical knowledge.

Keywords Mathematical practice · gaps in proofs · peer review in mathematics · the nature of proofs

1. Introduction

This paper examines the character of gaps in ordinary mathematical proofs, as opposed to formal proofs.[2] According to Don Fallis (2003), an ordinary mathematical proof is essentially a sequence of basic mathematical inferences where 'a basic mathematical inference' is an inference that is "accepted by the mathematical community as usable in proof without any further need of argument" (pp. 49–50). He states that the set of basic mathematical inferences varies across time and subdisciplines, but that it appears to be fixed in any given

---

[1] Line Edslev Andersen
lea@css.au.dk
Centre for Science Studies, Department of Mathematics, Aarhus University, Ny Munkegade 118, 8000 Aarhus C, Denmark

[2] Some mathematical proofs are formal proofs. A formal proof consists in a sequence of formulas in a formal language, each of which is either an axiom or the result of applying one of the explicitly stated rules of inference to previous formulas in the sequence. However, this paper is about gaps in ordinary mathematical proofs and ordinary mathematical proofs, as they occur in mathematical practice, are not formal proofs.

context and that "mathematicians know exactly what the set contains" (Fallis 2003, p. 49). A proof has a gap wherever it deviates from being a proof in this sense.

Fallis distinguishes between three types of gaps in ordinary proofs: inferential gaps, untraversed gaps, and enthymematic gaps. A mathematician has overlooked an inferential gap when the sequence of propositions she has in mind as being a proof is not a proof (Fallis 2003, p. 51). She has left an untraversed gap when she has not gone through all the details of the sequence she has in mind as being a proof (pp. 56–57). In the case where nobody else has gone through the details either, the gap is universally untraversed. Finally, a mathematician has left an enthymematic gap in the presentation of a proof when she has not presented the entire sequence of propositions she has in mind as being a proof (p. 53). Note that if she leaves a gap in the presentation of the proof, this qualifies as an enthymematic gap whether or not she has traversed the gap in her head (cf. Fallis 2003, p. 57). Untraversed gaps and enthymematic gaps are intentional gaps, and these are the focus of this paper.[3] To give an example, if someone has written in a proof that "A simple proof by induction shows that *p*" without giving the proof by induction, then she has left an enthymematic gap in the proof, a gap in the proof as written. If she has not gone through the proof by induction in her head either, she has left an untraversed gap in the proof. In the case where nobody else has gone through the details either, the gap is universally untraversed.

Fallis does not explicitly address the question of what types of gaps are left by whom.[4] The examples he gives in the sections on inferential and enthymematic gaps are about authors who leave gaps in their own proofs, but he implies that a mathematician can also leave inferential or enthymematic gaps in someone else's proof when she herself interprets or presents that proof. For example, she

---

[3] In a literature review on non-deductive methods in mathematics, Alan Baker (2015) has a section on gaps in proofs where he recounts Fallis' work in (2003). He notes that the notion of 'proof gap' is in need of further clarification (Baker 2015, section 2.1.2). Yacin Hamami (2014) takes steps in this direction by providing a more detailed account than Fallis of basic mathematical inferences. In particular, he applies Dag Prawitz's account of valid inference (e.g., Prawitz 2012) to mathematics. Hamami suggests that we think of Fallis' categories of gaps relative to the resulting account of valid mathematical inference. For our purposes, it is not necessary that we have a clear notion of what a complete proof is. An intuitive notion of this is enough.

[4] We thank a referee for suggesting that we address this question.

may commit a logical error when trying to fill in enthymematic gaps left by the author. She would then leave an inferential gap in the proof as interpreted by her. In the section on untraversed gaps, Fallis speaks about how both authors and readers of proofs leave untraversed gaps. He notes that mathematicians often leave untraversed gaps when reading published proofs, but that the authors of the proofs do not traverse all gaps either. In this paper, when we speak of untraversed gaps, we mainly focus on gaps that are left untraversed by mathematicians when they validate proofs as refererees, but we also address gaps left untraversed by readers in general when validating proofs by others. When we speak of enthymematic gaps, we focus on gaps authors leave in published proofs. More specifically, we focus on the type of enthymematic gaps that referees allow authors to leave in their proofs.

Fallis' (2003) main focus are the gaps that have not been traversed by anyone: universally untraversed gaps. His aim is to establish that there are proofs with universally untraversed gaps that are accepted by the mathematical community. He does not examine the general character of universally untraversed gaps or the other types of intentional gaps. The aim of this paper is to examine the character of intentional gaps in the case of published proofs. Unlike Fallis, we mainly focus on untraversed gaps and enthymematic gaps, as opposed to universally untraversed gaps. However, by saying something about the character of untraversed gaps and enthymematic gaps in published proofs, we also say something about the character of universally untraversed gaps in published proofs. In the case of enthymematic gaps, this is clear, since every universally untraversed gap is also an enthymematic gap. In the case of untraversed gaps, we aim to say something about the gaps mathematicians *typically* leave untraversed when validating proofs. By doing so we also say something about the character of universally untraversed gaps, although there are probably universally untraversed gaps that are not typically untraversed gaps (and vice versa). For it is fair to assume that there is a big overlap between the set of universally untraversed gaps and the set of typically untraversed gaps.[5]

Fallis makes his argument that there are proofs with universally untraversed gaps that are accepted by the mathematical community by giving *examples* of proofs that appear to have such gaps. Studying the general character of intentional gaps in proofs requires a different approach. This paper is thus based on in-depth interviews we conducted with mathematicians. We interviewed them about how

[5] We thank the two referees for pressing us to be more clear on the relationship between enthymematic gaps, untraversed gaps, and universally untraversed gaps.

they referee papers for mathematical journals. Examining mathematicians' refereeing practices can help us determine the kinds of proof gaps that are acceptable in published proofs. Our interviews suggest that a referee, when she checks a proof for correctness, also checks whether other experts would be able to check the proof for correctness. We focus on these two aspects of mathematicians' refereeing practices: on how referees check proofs for correctness and on how they check proofs for "checkability" by the relevant experts. As we shall see, the first aspect sheds light on the character of *untraversed gaps* in published proofs, while the second aspect sheds light on the character of *enthymematic gaps* in published proofs. These results in turn provide some new perspectives on the traditional philosophical questions of what grounds mathematical knowledge and of the nature of proof.[6]

After having introduced the interviews (section 2), we present the results on how referees proceed when checking a proof for correctness. These results tell us about the gaps that referees, and presumably mathematicians in general, leave untraversed when they validate proofs. It appears that certain subproofs are not validated through the checking of details, but by holding the subproofs, considered in broad outline, up against the landscape of mathematical knowledge (section 3). On this basis, we revise the traditional account of when a mathematician is justified in believing that a mathematical proposition p is true, on which she must have gone through the proof of p step by step (section 4). We then turn to the subject of enthymematic gaps and the results on how referees check a proof for checkability by the relevant experts. Referees appear to be guided by the criterion that enthymematic gaps are allowed when they can rather easily be traversed or be seen to be traversable by a large majority of the experts (section 5). On this basis, we examine how the 'right' amount of inferential rigor in published mathematical proofs is brought about by their dialogical character (section 6).

2. The interviews

Our account of mathematicians' refereeing practices is based on interviews we conducted with tenured researchers, of various academic ages, in different fields of pure mathematics. They are Danish and employed at Danish universities, and

---

[6] We are grateful to a referee for pressing us to develop further the philosophical motivations and consequences of the empirical study.

they are all male.[7] The choice of interviewing Danish mathematicians, in their and our first language, was made in order to allow the interviewees to speak as freely as possible and lower the risk of misunderstandings. We presume their nationality has not substantially influenced their answers. We interviewed an eighth mathematician, a Belgian, at a Belgian university and his responses did not stand out from the others. Danish universities were chosen for practical reasons. We have translated the quotations from the interviews used in this paper and also modified them to remove repetition and fillers. The quotations were subsequently reviewed by the interviewees, resulting in very minor modifications, and approved for publication.

The interviews had two parts. First, we asked the interviewees about specific experiences they had had with peer review in mathematics. We began by asking them to think about the last time they had refereed a paper for an international journal and we proceeded to ask them questions, one at a time, about this experience (how did you proceed; did you find any errors or shortcomings in the paper and, if so, of what kind). We then asked them to think about the last time they had a paper refereed by an international journal and asked them questions about this experience (how was the feedback you received; did you agree with the referee or referees), and so on. We asked them about specific experiences to keep them from philosophizing about mathematics and to try to get at their actual practice.[8] Although some of the interviewees quickly began to speak about other experiences they had had with refereeing papers and having papers refereed, they succeeded in mostly talking about specific experiences.

In the second half of the interviews, we asked more general questions, e.g. what counts as an error or shortcoming in a proof; can you say something about how detailed an argument has to be to constitute a proof; when you are trying to determine if a proof is valid, do you check every step of the proof (this question was taken from Weber 2008, p. 438). The interviewees were asked to respond to these questions on the basis of their general experience as journal referees. We ended both parts of the interview by asking questions about other topics than peer review – about differences in proof validation practices across mathematicians; about their reliance on others' results in their own work; and

[7] The number of female tenured mathematicians that are Danish and working at Danish universities is very small.

[8] We thank Henrik Kragh Sørensen and Mikkel Willum Johansen for suggesting this approach.

about errors in published proofs. The interviewees' responses to these questions are only used to a very limited extent here.

We transcribed the interviews, immersed ourselves in the transcripts, and developed codes grounded in the transcripts (as opposed to, e.g., codes grounded in the literature). The final coding taxonomy comprises eight themes (the level of thoroughness of the referee; the role of the background knowledge of the referee; believability of the results in the proof; believability of the approaches used in the proof; line by line checking of the proof; checking for checkability of the proof; proof errors; adding details). We had originally planned to use the interviews as a basis for an account of differences in how mathematicians referee papers, but when transcribing and rereading the interviewees we came to see a different story emerge about similarities in approach across mathematicians. It is the similarities in approach suggested by the interviews we present here. These results are obviously very tentative because of the small sample.

We should mention that papers submitted to mathematical journals are usually only reviewed by a single referee (Geist et al. 2010, p. 161; Grcar 2013, p. 422). This was confirmed by our interviews, although some of the interviewees mentioned that using more than one referee *is* the standard practice of some of the very top-ranked journals.

3. Proof validation and untraversed gaps

When a referee checks a proof for correctness, she apparently begins by holding the proof, considered in broad outline, up against the landscape of what she knows. Interviewee 4 explained that he begins by trying to get an overview of what the article says, to "form an impression of whether I believe it. Is it consistent with what I know or is there something that jumps out at me? For example," he continued, "I recently refereed an article in which there was an example that was inconsistent with something in the literature." Importantly, this approach is not only used to find mistakes in proofs, but also to validate proofs. Interviewee 5 said that,

> When you have been in this profession for a long time, you have gained mathematical insights that sometimes enable you to quickly see if a new proof or part of a proof within the scope of your interest is generally correct or not. If you don't ascend to that perspective – put on those

> glasses – but try to elbow your way through it, then you may calculate, calculate and calculate, and eventually come to see if it is correct or not.

Referring to the same experience-based ability, Interviewee 7 noted that "I only become convinced that something is a proof when there's nothing that stinks! You can smell it when something is not right." It is relevant to mention here Eva Müller-Hill's interviews with six mathematicians. Two of the interviewees mentioned that referees are not actually expected to go back to "step one" when checking a proof, but only to check the parts that look "suspicious" (Müller-Hill 2011, pp. 307–308, pp. 327–328; the word 'suspicious' was used by both).

The interviews further suggest that the referee does two types of thing when she checks the proof in broad outline against what she knows: She checks whether each subresult of the proof seems reasonable in light of what she knows and, at least for most of the subresults, whether it seems reasonable that this type of result can be proved in this type of way, with this type of tools. If so, she will usually not go on to check the subproof line by line. We may thus speak of two types of proof validation. We call them Type 1 validation, or validation by comparison, and Type 2 validation, or line-by-line validation. Since the subproofs that the referee validates using Type 1 validation are not validated through the checking of details, the referee will leave relatively many untraversed gaps in these.

For example, Interviewee 1 brought up mathematicians' ability to evaluate whether a new result is true independently of its claimed proof – which is needed for Type 1 validation – by saying that, "when you meet a result in your own field, you have some sense of whether it is right or wrong." He explained that when you sense that a result is correct and someone then claims to have proved it, she sounds convincing. Interviewee 1 and 2 described being extra careful when a result is "surprising."

The focus on the tools used to prove results implied by Type 1 validation was emphasized by Interviewee 6 in the following way: He stated that "I look at the result a lot and try to think of similar propositions that I have seen earlier. I ponder whether I believe that the ingredients used by the author should be strong enough to prove the result in question." This is similar to a point made by mathematician William Thurston. "When I read a mathematical paper in a field in which I'm conversant," he wrote, "I might look over several paragraphs or strings of equations and think to myself, 'Oh yeah, they're putting in enough rigamarole to carry such-and-such idea'" (Thurston 1994, p. 167). Earlier in the interview, Interviewee 6 had looked back on the referee reports he had himself received over the years and described how there had been three common types of comments. Describing the third type, he said that, "[Sometimes] a referee will write something like, 'you prove this result, but the techniques you use should be able to prove a stronger result'." Sometimes he agrees with the referee. Other times, he said, "I have spent months trying to convince myself that the referee is wrong so that I can write a short comment to him, explaining why the stronger results cannot be proven by the same techniques." This illustrates in a different way how referees compare the claimed results with the tools used when validating proofs.

Type 1 validation can be quite strong in the sense that referees by using this approach may become convinced of the general validity of a proof in spite of not fully understanding the proof or in spite of minor errors in the proof. Interviewee 1 stated that,

> Maybe there are some minor things in a proof that I sense are okay but where I don't quite understand the details. I feel that the reason for my lack of understanding is my own ignorance because the proof looks pretty standard and I have seen other proofs that resemble it. […] Then I take it to be okay.

Interviewee 3 noted that, "If I can see that a proof will work, but that there must be [an insignificant calculation error] somewhere […], then I may ignore this and not go through the calculations to try and find the error."

Checking a proof in broad outline against what the referee knows is sometimes not an immediate option, since the referee may not at all be able to see "what is going on" in some subproof. Examples can often help her with this. For example, Interviewee 2 said that,

> When I cannot immediately follow the chain of reasoning from A to B, I will use examples. I will typically start by choosing the simplest example I can think of and try to see whether the proof works in this case. If the answer is yes, then I might choose a more complex example, and then sometimes I come to realize what is going on from A to B. It also happens, of course, that it turns out not to be correct that B follows from A. Maybe I have found an example in which it does not hold.

When he in this way comes to see what is going on from A to B, he sometimes proceeds to do a line by line checking of the subproof and sometimes not. This presumably depends on how well the claimed path from A to B fits into what he knows.

When the referee has checked the proof, in broad outline, against what she knows, she turns her focus to the parts that stand out as surprising or suspicious. The referee will typically check these more or less line by line and leave few or no untraversed gaps. When checking a subproof line by line, the referee will often find that the author has left too many *enthymematic gaps*, that the author has not provided enough detail for her to be able to follow every step of the proof as stated. Sometimes the referee will be able to supply the needed details herself. Sometimes she will not be able, in the limited time she has available, to provide the extra details. She will then typically remark on this in the referee report, writing something along the lines of, "I cannot see how you get from A to B, please provide more detail." This leads to the eliminiation of enthymematic gaps in the proof. But what enthymematic gaps are acceptable in published proofs is really determined by something else, which is why we do not focus on enthymematic gaps in the present section. As mentioned in the introduction, and as will be described in detail in section 5, referees apparently follow the criterion that the gaps in the published proof should be rather easily traversable or seen to be traversable by a large majority of the relevant experts. This checkability criterion will generally place a stricter limit on enthymematic gaps than the expertise of the referee. In the rare case where the referee is less of an expert with respect to the topic of the proof than a large majority of the mathematicians in the specialized field, her limited expert experience will lay down a stricter limit than the checkability criterion. What makes this case particularly rare is that the journal editor will typically pick a referee that has published on a topic closely related to the topic of the submitted proof (Geist et al. 2010, pp. 160–161).

Note that the above description of the refereeing process could help explain why published proofs often contain minor errors, but rarely critical or unrepairable errors (Davis 1972, pp. 260–262). These two characteristics of published proofs were mentioned by just over half of the interviewees when we brought up the topic of how mathematicians go about correcting errors in published proofs. That not every step of the proof is checked by the referee can help explain the abundance of minor errors. That the referee has checked the proof in broad outline against what she knows can help explain the lack of critical errors. This may in turn be part of the reason why mathematicians rarely publish errata (Grcar 2013).

4. Consequences for the Cartesian story

Fallis' paper on gaps addresses the question of when a mathematician is justified in believing that a mathematical proposition *p* is true. According to the traditional "Cartesian story" (Fallis 2003, pp. 46–47), she must have gone through the proof of *p* step by step. Fallis argues that this story fails as a descriptive account of when she is justified *in a manner that is accepted by the mathematical community* in believing that *p*; a mathematician can be justified, in this manner, in believing that *p* when there are gaps in the proof that neither she nor anybody else has traversed. It is in this context that he gives examples of proofs that are accepted by the mathematical community and that appear to have gaps in them that have not been traversed by anyone.

Another reason why the Cartesian account is wrong is apparently that a mathematician can be justified in believing that *p* through testimony (Hardwig 1991). We set that aside here and, like Fallis, focus on the question of when a mathematician is justified in believing that *p* from knowing a proof of *p*.

If section 3 accurately describes how mathematicians typically validate proofs when acting as referees, it seems reasonable to suppose that mathematicians, whether or not they are acting as referees, typically validate proofs in this type of way.[9] This means that they leave untraversed gaps in the proofs in either case. If this is true, a mathematician does not have to go through a proof of *p* step by step to be justified in a manner that is accepted by the mathematical community in believing that *p*, as Fallis also argues. Fallis thus ends his paper by stating that "the Cartesian story must be supplemented in some way" (2003, p. 64). The interviews suggest that we begin by supplementing the story with this: A mathematician can be justified in a manner that is accepted by the community in believing that *p* when she has validated a proof of *p*, although parts of the proof have been validated using Type 1 validation. Knowledge of other proofs, of how certain types of results can be proved in certain types of ways, can thus play a significant role in a mathematician's being justified in believing that *p* from knowing a proof of *p*.

This is relevant to the debate between formalists and their opponents on why mathematicians believe theorems. While the former hold that formal derivations underlying ordinary proofs, although usually not executed, are the reason why

[9] Mathematicians may on average be less thorough when they are not acting as referees, since the interviews suggest that referees take the task of determining whether the submitted paper is sound very seriously (Andersen 2017).

mathematicians believe theorems, the latter reject or significantly weaken the role of formal derivations in the gaining of confidence in theorems (Pelc 2009, pp. 84–87). The discussion between Azzouni and Rav (Rav 1999; Azzouni 2004; Rav 2007) prominently exemplifies the debate. Azzouni writes that, "Ordinary mathematical proofs *indicate* (one or another) mechanically checkable derivation of theorems from the assumptions [of] those ordinary mathematical proofs" (2004, p. 105; emphasis in original). He defends the view that mathematicians essentially validate a proof by recognizing a mechanically checkable derivation indicated by the proof, which can be done without writing the derivation out. Some philosophers have convincingly argued against this view on a priori grounds, including Azzouni himself, who later changed his view (e.g., Azzouni 2009; Pelc 2009; Tanswell 2015). My interviews appear to provide empirical grounds against it. For a mathematician to recognize the particular derivation claimed to be indicated by a proof, all the details of the proof should be important, but the interviews suggest that they are not all important in validating the proof; when a subproof is validated using validation by comparison, many details are unimportant. In addition, none of the interviewees brought up the concept of formalizability in any way.

Müller-Hill (2009) has also written about formalizability and knowledge ascriptions in mathematics. She writes on the assumption that the following criterion is a good criterion for knowledge ascription in mathematics: (*) S knows that *p* if and only if S has available a formalizable proof of *p* (Müller-Hill 2009, p. 21). Based on a questionnaire study of 76 mathematicians, she examines how the criterion must be interpreted to accord with how mathematicians ascribe knowledge to mathematicians. Later she conducted interviews with six mathematicians about the same subject. One of the questions she asked them was whether they believe that the proofs that are accepted by the mathematical community are formalizable (Müller-Hill 2011, pp. 305–346). They answered in the affirmative with the qualification "in principle." This is an interesting result on its own. It provides some evidence that mathematicians believe that (*) is correct and is evidence that proofs are only acceptable to the community if they are believed to be formalizable in principle. But these data do not by themselves tell us about what a mathematician must do to come to know that *p* from studying its proof, which is the question we focus on here. Mathematicians may believe that the proofs that are accepted by the mathematical community are formalizable without this really playing a role in ordinary proof validation. Müller-Hill's interview transcripts leave open the possibility that the concept of formalizability plays no important role in ordinary proof validation and are thus consistent with the results of our interviews.

5. Proof checkability and enthymematics gaps

The interviews suggest that a referee checks a proof for "checkability" by the relevant experts, so when she checks the proof for correctness in the way described above, she also takes into account whether others would be able to check the proof for correctness in the same type of way. Hence, she does not only comment on the parts that she cannot follow herself, but also on parts that she believes will be hard for others to follow unless more information is provided. In doing so, she appears to be guided by the criterion that a large majority of the experts in the specialized field should be able to validate the proof within a reasonable amount of time.[10] This criterion does not ask for the expert to be able to traverse any gap within a reasonable amount of time, but for her to be able to see that she could do so if she tried. The authors often, if not almost always, submit proofs that have unacceptable enthymematic gaps in them, presumably in part because of their being fully immersed in the topic. The journal referee represents the group of relevant non-immersed experts and aims to eliminate these gaps or to ensure that the author eliminates them.

When Interviewee 3 told us about discussions he had with a collaborator about the appropriate level of detail in their joint work, he said that, "A proof should preferably be so detailed that maybe 80 percent of the people you meet at the specialized conferences can read and understand it." He added, "In general I think you should strive to write a little bit more than what you yourself think necessary; since you are so immersed in the topic, you would otherwise presumably write a bit too little." It is apparently not generally required that new PhD students should be able to read a proof within a reasonable amount of time. Interviewee 6 said that, "When I was a PhD student I thought, 'Why do I have to sit here and struggle to understand these steps – why couldn't the author just have written the argument out more!'"

The checkability criterion plays different roles in the two types of validation. In the subproofs that are validated using Type 1 validation, or validation by

---

[10] Kenny Easwaran (2009) offers an account of why mathematicians are generally unwilling to accept probabilistic proofs, but do accept proofs that skip steps and are long and complicated. (A probabilistic proof does not deductively establish its conclusion but establishes that there is some, often specifiable, high probability of the conclusion being true.) In a footnote, he states, referring to enthymematic gaps, that he thinks that "the sorts of proof gaps that are acceptable are the ones that relevant experts can see and still be convinced" by the proof (Easwaran 2009, p. 355). My interviews support this picture.

comparison, the criterion does not lead to the elimination of enthymematic gaps. When the referee validates a subproof by checking it in broad outline against what she knows, but does not believe that most of the other experts would be able to validate it in this way, it is unlikely that she will check the subproof line by line to ensure that they can become convinced in this way instead, especially given the limited time she has available. Rather, the referee will try to ensure that the paper enables them to check the subproof using validation by comparison. She can do so by asking the author to provide an introduction that describes the context of the part of the proof in question or to give more references to the literature in other ways. When we asked Interviewee 4 about the last referee report he had received, he said,

> The referee thought the context needed to be explained better. If the article had addressed a small group of experts, then they wouldn't have bothered to read such explanation. They would know it already. But the article addressed two groups of experts, and you could say that both groups needed a bit more explanation; the experts in one group were not experts in the area of the other group and vice versa.

When speaking in general terms about the feedback he had received from referees, he said, "Sometimes the referee will say that there is a lack of references," among other things, "so the referee has probably done something similar to what I would have done: tried to place the proof in its context."

When the referee moves on to validate parts of the proof line by line, she will comment on the steps that she believes will be hard for experts to follow unless more details are provided. Thus, the checkability criterion does lead to the elimination of enthymematic gaps in the parts of the proof that are validated using Type 2 validation, line-by-line validation. For example, Interviewee 3 said that, "If I have thought for five minutes about one step and still cannot see what is going on, then it is probably a good idea to ask the authors to add an extra step in order to improve the readability." When we asked Interviewee 2 about the last referee report he received, he told us that the referee had suggested that details be added in certain places. In explaining why he had chosen to follow this suggestion he said that, "When you have written an article, there are many things that you take to be obvious, and the referee may have had a point in saying that this here is not as obvious to the reader as it may be to you." Later, when we were asking general questions, we asked him about what counts as an error or a shortcoming in a proof. He responded that,

> There will be an error or a shortcoming in a proof if there is a lot for the reader to fill in, in order to get from A to B. […] As a referee, I have sometimes asked the author to fill in some details at a certain place simply because I haven't found it reasonable that the reader should do so himself, or because I haven't myself been able to see what the argument really was.

In this way, the refereeing process limits the amount of enthymematic gaps in the surprising or suspicious parts of published proofs.[11] We do not mean to say that this holds for every proof that has been subjected to peer review. It happens that a referee is very unthorough. Still, if referees typically proceed as described above, the refereeing process significantly limits the number of enthymematic gaps in the surprising parts of published proofs.

Note that if an editor discovers that a referee has been unthorough, she may invite someone else to review the paper, but we do not know to what extent this is common practice. In general, we know very little about editor practices in mathematics. Geist, Löwe, and Van Kerkhove (2010) describe the results of a small questionnaire about the peer review process that they sent to editors of mathematical journals. Among other things, respondents were asked to pick one of these statements: "I think the referee should check a. all/b. some/c. none of the proofs in detail." Six respondents selected option (a) and five respondents selected option (b). Option (c) was not selected by anyone. Respondents were also asked the following question: "What percentage of referees approximately do a good job checking the correctness of a paper's claim?" The average of the answers was 52.3 percent. Thus, the responding editors believe referees should check all or some of the proofs in the submitted paper in detail and often experience that the referees do not do this. But we do not know how the editors react to this (e.g., to what extent they then ask someone else to referee the paper) or the level of their dissatisfaction. We also do not know to what extent the six editors that picked option (a) would be happy with referees that proceed as described above.

The checkability criterion primarily tells us about enthymematic gaps, but by limiting the number of enthymematic gaps in published proofs, it also limits the number of untraversed gaps in published proofs; it limits the extent to which the referee and the author can leave untraversed gaps in the sequence of propositions they have in mind as being a proof. Although the referee must

---

[11] At the same time, a referee may very well ask an author to provide less detail – i.e. to leave *more* enthymematic gaps – in the straightforward parts of a proof. We are grateful to a referee for pressing us to clarify our claim here.

often ask the author to provide more details in the proof as written, the author will often have gone through more details of the proof in her head than the referee has. Our point is just that there is a strict, discernible limit on how lax in their thinking the author and the referee can be when the referee proceeds as described above.

6. Consequences for the nature of proof

Contrary to what is commonly thought, A. C. Paseau (2016) argues that there is no epistemic value such that completely inferentially rigorous arguments have that value in virtue of being completely inferentially rigorous. Paseau's paper concerns arguments in general, with a particular focus on mathematical proofs. When he speaks of completely rigorous mathematical proofs, he is not referring to formal proofs but to gapless proofs in Fallis' sense, although he prefers to call them atomized proofs (Paseau 2016, pp. 178–181). Paseau writes that, "A positive formulation of our main point is that the 'right' amount of inferential decomposition is epistemically valuable, but that usually that right amount falls far short of atomization" (2016, p. 187). What the right amount is varies, of course, with the considered epistemic value. Our study suggests what the relevant epistemic value is when we are talking about proofs in mathematical journals, namely checkability by the experts, and how the right level of inferential decomposition falls short of atomization in this case. Given this value, the right level of decomposition depends on the relevant subcommunity and also varies across different parts of the proof, depending on whether they can be validated using validation by comparison.[12]

As we have seen, the referee, who is not immersed in the topic of the proof like the author is, plays an important role in ensuring that the proof has "the 'right' amount of inferential decomposition." The referee represents in a strong sense

[12] This is a very partial response to the following remarks by Grice: "Finicky over-elaboration of intervening steps is frowned upon, and in extreme cases runs the risk of forfeiting the title of reasoning. In speech, such over-elaboration would offend against conversational maxims, against (presumably) some suitably formulated maxim of Quantity. In thought, it will be branded as pedantry or neurotic caution. At first sight, perhaps, one would have been inclined to say that greater rather than lesser explicitness the better. But now it looks as if proper explicitness is an Aristotelian mean, and it would be good some time to enquire what determines where that mean lies" (Grice 2001, p. 16; quoted in Paseau 2016, p. 187).

the subcommunity of relevant experts; by checking the proof for checkability by the experts she speaks with the voice of the subcommunity. An experienced author will presumably have written and checked her proof with this voice in the back of her mind. During the refereeing process, she is directly confronted with it and will try to accommodate it. Contained in the published proof are the author's responses to the responses of this voice to the author. In this sense, the proof can be said to consist in a dialogue between the relevant community and the author.

The interviews thus provide evidence for Catarina Dutilh Novaes' (2016, 2017) account of proofs as dialogues between what she calls Prover and Skeptic. On her account,

> A deductive proof corresponds to a dialogue between the person wishing to establish the conclusion […] and an interlocutor who will not be easily convinced and will bring up objections, counterexamples, and requests for further clarification. A good proof is one that convinces a fair but 'tough' opponent. Now if this is right, then mathematical proof is an inherently dialogical, multi-agent notion, given that it is essentially a piece of discourse aimed at a putative audience (Ernest 1994). (Dutilh Novaes 2016, p. 2617)

Later she writes: "Ultimately, most of the work is done by Prover, but Skeptic has an important role to play, namely to ensure that the proof is persuasive, perspicuous, and valid" (Dutilh Noaves 2016, p. 2618).

Dutilh Novaes' description of the role of Skeptic fits very well with the role of the referee as described above. The referee may herself be easily convinced of the validity of the proof, but as a representative of the experts she will not be easily convinced; she must become convinced that she would also be convinced of the validity of the proof if she were one of the other experts. This supports Dutilh Novaes' emphasis on the audience and suggests that we speak, as we do above, of a published mathematical proof as a piece of discourse *with* the relevant audience, not only as a piece of discourse aimed at that audience. It seems that we may speak in this way of mathematical proofs in general. Other presentations of proofs, such as a blackboard presentation of a proof to a colleague or at a conference, also seem to be pieces of discourse with audiences, in virtue of the speaker preparing the proof for presentation with the particular audience in mind and making changes to the proof in response to questions and comments from the colleague or the conference attendees. In these cases, the amount of inferential decomposition will be smaller because the audiences are smaller, but mainly because of the oral format (for an account of the difference in style between oral and written proof presentations, see Johansen and Misfeldt 2016). Consequently, while Dutilh Novaes conceptualizes mathematical proof as a dialogue between Prover and Skeptic, whom she describes as fictitious (2016, p. 2618), we would rather conceptualize it as a dialogue between the author and an actual audience (be that a mathematical subcommunity as represented by a referee, the participants in a conference, or a colleague).

7. Conclusion

Through mathematicians' refereeing practices we have examined the character of untraversed gaps and enthymematic gaps in published proofs. We found that referees use two methods of validation when reading proofs and that this is important to when these gaps are allowed in published proofs. Relatively many gaps are allowed in the parts of a proof that the referee can see and help others see fit right into the literature, since these parts are not validated through the checking of details, but by using what we have called Type 1 validation. By contrast, relatively few untraversed gaps and enthymematic gaps are allowed in the parts of the proof that the referee validates by checking them line by line. When subparts of these describe standard moves, relatively many gaps are again allowed. We have used these results to contribute to traditional discussions in the philosophy of mathematics. We have thus argued that a mathematician can become justified in believing that *p* through Type 1 reasoning and, like Dutilh Novaes (2016), that we should conceive of proofs as dialogues that provide an appropriate level of rigor.

So far, we have not been concerned with the type of gaps that is Fallis' main concern: universally untraversed gaps. We end by addressing these. As mentioned in section 5, the checkability criterion limits the number of untraversed gaps in published proofs by limiting the number of enthymematic gaps in published proofs. The checkability criterion thus also limits the number of *universally* untraversed gaps in published proofs. It is also worth noting that a proof will often continue to be checked after it has been published. The mere use of a result by others works as a checking mechanism. For example, a mathematician often does not have the option of using others' results without studying their proofs. Some of the interviewees emphasized this. Interviewee 1 stated that, "Maybe you want to fine-tune or generalize some of the arguments in the proof of the result, or something like that. In that case, you are automatically thrown into the proof. Sometimes you discover problems in the proof, sometimes not." This type of use of a result would likely require you to go through the proof very thoroughly. Also, if the use of the result leads to surprising or unreasonable results, mathematicians will read the proof again in search for errors. Some of the interviewees brought up this way of discovering errors in proofs (so does Devlin 2003). Hence, informal post-publication peer review is likely to further decrease the amout of universally untraversed gaps in the proof. When there are gaps in the proof that none of the readers have traversed, these mathematicians have at least partially independent evidence, based on their expert experience, that the gaps could be traversed if they tried (cf. Fallis 2003, p. 62; Paseau 2011, p. 145). This account suggests that universally untraversed gaps in published proofs are, while not necessarily few, quite innocent.

Acknowledgements

Part of the research for this paper was conducted while I was a postdoc at the Centre for Logic and Philosophy of Science, Vrije Universiteit Brussel, Brussels, Belgium. Two years ago, I had the pleasure of interviewing eight mathematicians about their refereeing practices. I thank them for their time and openness. The paper has benefited greatly from comments from Henrik Kragh Sørensen, Karen Francois, Mikkel Willum Johansen, and two anonymous referees. I also thank Henrik and Mikkel for their help in developing the interview questions. Earlier versions of the paper were presented at the 2016 Society for Philosophy of Science in Practice conference (Glassboro, New Jersey), the 2016 Novembertagung on the history and philosophy of mathematics (Sønderborg, Denmark), the workshop 'Mathematical evidence and argument: Historical, philosophical, and educational perspectives' (Bremen, Germany, 2017), the workshop 'Group knowledge and mathematical collaboration' workshop (Oxford, UK, 2017), and the 2017 Nordic Network for Philosophy of Science meeting (Copenhagen, Denmark). I would like to thank the audiences for useful comments.

Conflict of interest

The author declares that she has no conflict of interest.

References

Andersen, L. E. (2017). On the nature and role of peer review in mathematics. *Accountability in Research, 24,* 177–192.

Azzouni, J. (2004). The derivation-indicator view of mathematical practice. *Philosophia Mathematica, 12,* 81–105.

Azzouni, J. (2009). Why do informal proofs conform to formal norms? *Foundations of Science, 14,* 9–26.

Baker, A. (2015). Non-deductive methods in mathematics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/fall2015/entries/mathematics-nondeductive/. Accessed 14 December 2016.

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

Devlin, K. (June 2003). When is a proof? In *Devlin's angle*, a column written for the Mathematical Association of America. https://www.maa.org/external_archive/devlin/devlin_06_03.html. Accessed 14 December 2016.

Dutilh Novaes, C. (2016). Reductio ad absurdum from a dialogical perspective. *Philosophical Studies, 173,* 2605–2628.

Dutilh Novaes, C. (2017). A dialogical conception of explanation in mathematical proofs. Forthcoming.

Easwaran, K. (2009). Probabilistic proofs and transferability. *Philosophia Mathematica, 17,* 341–362.

Ernest, P. (1994). The dialogical nature of mathematics. In P. Ernest (Ed.), *Mathematics, education and philosophy: An international perspective* (pp. 33–48). London: The Falmer Press.

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

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

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

Grice, H. P. (2001). *Aspects of Reason*. Oxford: Oxford University Press.

Hamami, Y. (2014). Mathematical rigor, proof gap and the validity of mathematical inference. *Philosophia Scientiæ, 18,* 7–26.

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

Johansen, M. W., & Misfeldt, M. (2016). An empirical approach to the mathematical values of problem choice and argumentation. In B. Larvor (Ed.), *Mathematical cultures: The London meetings 2012–2014* (pp. 259–269). Basel: Birkhäuser.

Müller-Hill, E. (2009). Formalizability and knowledge ascriptions in mathematical practice. *Philosophia Scientiæ, 13,* 21–43.

Müller-Hill, E. (2011). Die epistemische Rolle formalisierbarer mathematischer Beweise. Formalisierbarkeitsorientierte Konzeptionen mathematischen Wissens und mathematischer Rechtfertigung innerhalb einer sozio-empirisch informierten Erkenntnistheorie der Mathematik. Inaugural-Dissertation. Bonn: Rheinische Friedrich-Wilhelms-Universität. http://hss.ulb.uni-bonn.de/2011/2526/2526.htm. Accessed 14 December 2016.

Paseau, A. C. (2011). Mathematical instrumentalism, Gödel's theorem, and inductive evidence. *Studies in History and Philosophy of Science, 42,* 140–149.

Paseau, A. C. (2016). What's the point of complete rigour? *Mind, 125,* 177–207.

Pelc, A. (2009). Why do we believe theorems? *Philosophia Mathematica, 17,* 84–94.

Prawitz, D. (2012). The epistemic significance of valid inference. *Synthese, 187,* 887–898.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica, 7,* 5–41.

Rav, Y. (2007). A critique of a formalist-mechanist version of the justification of arguments in mathematicians' proof practices. *Philosophia Mathematica, 15,* 291–320.

Tanswell, F. (2015). A problem with the dependence of informal proofs on formal proofs. *Philosophia Mathematica, 23,* 295–310

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

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