On mathematicians' disagreements on what constitutes a proof

Authors: Keith Weber and Jennifer A. Czocher

Contact and affiliation information:

Keith Weber

Rutgers University

Graduate School of Education

10 Seminary Place

New Brunswick, NJ 08901

U.S.A.

keith.weber@gse.rutgers.edu

Jennifer Czocher

Texas State University

Department of Mathematics

San Marcos, TX 78666

U.S.A.

czocher.1@txstate.edu

On mathematicians' disagreements on what constitutes a proof

Abstract. We report the results of a study in which we asked 94 mathematicians to evaluate whether five arguments qualified as proofs. We found that mathematicians disagreed as to whether a visual argument and a computer-assisted argument qualified as proofs, but they viewed these proofs as atypical. The mathematicians were also aware that many other mathematicians might not share their judgment and viewed their own judgment as contextual. For typical proofs using standard inferential methods, there was a strong consensus amongst the mathematicians that these proofs were valid. An instructional consequence is that for the standard inferential methods covered in introductory proof courses, we should have the instructional goal that students appreciate why these inferential methods are valid. However, for controversial inferential methods such as visual inferences, students should understand why mathematicians have not reached a consensus on their validity.

Keywords: Agreement; Mathematicians; Proof

1 Introduction

Because proving is regarded as central to mathematicians' practice, there is a consensus among mathematics educators[1] that proving should play a central role in all mathematics classrooms (Stylianides, Bieda, & Morselli, 2016; Stylianides, Stylianides,

[1] Throughout the paper, the term "mathematics educators" refers to researchers in mathematics education.

& Weber, 2017). Further, many mathematics educators maintain that the way that proof is practiced in mathematics classrooms should be informed by, and compatible with, the way that proof is practiced by the mathematical community (e.g., Dawkins & Weber, 2017; Harel & Sowder, 2007; Herbst & Balacheff, 2009; Stylianides, 2007), where the mathematical community is usually assumed to be comprised of research-active mathematicians who work at universities[2]. Mathematics educators do not expect mathematical classrooms to be exact replicas of professional mathematical communities with regard to proving, both because mathematicians have conceptual and representational resources that students lack (Stylianides, 2007; Weber, Inglis, & Mejia-Ramos, 2014) and because the needs of the mathematical community and the classroom community differ (Staples, Bartlo, & Thanheiser, 2012). Nonetheless, many mathematics educators expect students to develop standards of conviction and perceptions of proof that align with those held by professional mathematicians (Harel & Sowder, 2007) and desire students' proofs to meet some of mathematicians' professional standards (adapted to the needs and background of the classroom community) (Dawkins & Weber, 2017). At a minimum, teachers and mathematics educators have an obligation to present proving in a manner that does not distort, or is incompatible with, how mathematicians engage in this activity (Herbst & Balacheff, 2009).

Although many mathematics educators agree that classroom communities should engage in proving practices that are compatible with professional mathematical practice,

[2] It is not necessarily the case that the mathematical community *should* be defined in this way (Stillman, Brown, & Czocher, in preparation). We only observe that this is how the mathematical community is usually operationalized in mathematics education studies, in which participants are recruited from mathematics staff and the opinions of research-active mathematicians are cited. We also do not claim that the mathematical community that we described is homogeneous. Indeed, one purpose of this paper is to highlight their heterogeneity.

there are deep disagreements among mathematics educators as to what mathematicians' proving practices are or even what constitutes a proof to a mathematician. These disagreements have led to inconsistent recommendations on how proof should be taught and an incoherent literature base containing contradictory claims (Cirillo et al., 2015; Balacheff, 2008; Reid & Knipping, 2010)[3].

In this paper, we begin by raising a contentious question about mathematical practice: Do mathematicians agree on what constitutes a proof? We discuss how the answer to this question has important consequences for pedagogy and research; we further document that mathematics educators answer this question in different ways. Next, we introduce data from an empirical study illustrating the extent to which mathematicians agree and disagree on controversial proofs. We then use this data to bridge the contrasting positions and propose a resolution to the question. Finally, we explore the implications of our resolution for teaching mathematics, conducting mathematics education research, and understanding how mathematicians practice their craft.

2 Theoretical perspective

2.1 On empirical studies of mathematical practice

This paper is based on three premises about mathematics education research: (1) Mathematics educators' beliefs about proof should be informed by how mathematicians practice their craft. (2) To understand how mathematicians practice their craft, it is useful

[3] For a notable example, see the well-known debates about the relationship between argumentation and proof (Balacheff, 1999; Boero, 1999; Duval, 1999). These debates hinged on the extent that a proof needs to be a structured argument highlighting logical dependency within an axiomatic system (c.f., Balacheff, 1999; Mariotti, 2006), a matter that is still being disputed today.

to consider the work of scholars whose research concerns mathematicians' practices regarding proof. These scholars include philosophers as well as mathematicians who write reflective essays on their practice. (3) When the aforementioned scholars are unable to reach a consensus on some issue about mathematical practice, systematic empirical studies can provide useful data to inform these debates.

A full discussion of these theoretical premises is beyond the scope of this paper, but we briefly elaborate on each premise below. Regarding (1), some mathematicians (and philosophers in the analytic tradition) have defined proof syntactically—as a sequence of statements in a logical theory with explicit and precise rules for what constitutes permissible statements, axioms, and rules of inference. Mathematics educators nearly uniformly agree that this formal perspective on proof is inappropriate for forming the basis of mathematics instruction, both because few mathematicians actually produce proofs of this type and because adopting this perspective would focus students' attention on form over content. For further discussion, see CadwalladerOlsker (2011). Regarding (2), there is a rich tradition in mathematics education to appeal to philosophers of mathematics to gain insight into mathematicians' practice with regard to proof. For instance, some appeal to philosophers such as Rav (1999) to highlight the social nature of proof. Others appeal to Steiner (1978) to emphasize that proofs can have explanatory value. Still others appeal to Lakatos (1976) to note that theorems and proofs are corrigible. For further discussion, see Weber and Dawkins (2018). Regarding (3), sometimes scholars reach different conclusions about some aspect of mathematical practice. When this occurs, we adopt the simple assumption that mathematical practice is an empirical phenomenon and therefore how members of the mathematics community behave is subject to systematic empirical investigation. For further discussion, see Inglis and Aberdein (2016).

2.2 On the nature of proof

In this paper, we follow Stylianides (2007) in defining a proof as an argument that begins with true statements that are acceptable to a mathematical community, deduces new statements via valid logical inference, and is couched within an appropriate representation system[4]. Our paper focuses on what inference methods mathematicians accept as valid and appropriate in a proof. That is, following Balacheff (1999), when does the prover have the "license to infer"?

The goal of this paper is to investigate the question about the extent to which mathematicians agree on which inferential schemes are permissible. In the remainder of this section, we elaborate on two contrasting positions on the nature of mathematical proof. The first position, which we call the consensus view on proof, asserts that mathematicians usually or always agree on whether an argument qualifies as a proof (and consequently on which inferential schemes are permissible in a proof). The second position, which we call the pluralistic view on proof, asserts that different mathematical communities (and perhaps different mathematicians within the same community) disagree on what qualifies as a proof. This disagreement is due to mathematicians having different positions on what inferential schemes are permissible in a proof.

[4] Stylianides tailored his definition to be appropriate to classrooms and included that the accepted statements should be known by the classroom community and the inferential schemes should be accepted by, or within the conceptual grasp, of the community. In this paper, we only discuss statements and inferential methods that are within the conceptual grasp of the professional mathematical community, so these important nuances for classroom proofs will not be relevant here.

2.3 Sources of disagreement

To clarify the consensus and pluralistic views, it is useful to distinguish among three ways in which mathematicians may disagree on whether a particular argument is a proof. We classify these three types of disagreement as arising from performance errors, gap sizes, and permissible inferential schemes. The first source of disagreement is rooted in the fact that mathematicians sometimes overlook mistakes. That is, two mathematicians may disagree on whether an argument qualifies as a proof because one identifies an invalid step that the other has overlooked. We refer to this type of disagreement as resulting from *performance error*. A mathematician who accepted a proof because she failed to locate a flaw in the proof simply made an error. Those who hold a consensus view on proof usually do not deny that performance errors occur. Instead, they claim that disagreements due to performance errors can be resolved through discussion. For instance, Selden and Selden (2003), who adopted a consensus view, wrote that, "if two mathematicians disagree on the correctness of a proof, they will often attempt a joint validation of it (or fragment of it). Typically they will either expand the proof and agree on the expanded version or, failing that, they will find and agree on a mistake that cannot be fixed" (p. 7).

A second source of disagreement on whether an argument is a proof is due to different views on *whether a particular gap in a proof is too large*. In both mathematical practice and pedagogical practice, proofs commonly contain gaps in which routine logical steps are omitted. When a mathematician encounters a gap while she is evaluating the validity of a proof, she must estimate whether a knowledgeable audience member would have the capacity to bridge this gap with a series of logical inferences. The appropriateness of a gap therefore may be contextual in two respects. First, the permissibility of a gap may depend on the *purpose* of the proof. Perhaps mathematicians would be less tolerant of a gap if the goal of the proof was to rigorously establish a theorem for the first time in a research journal than if the same proof appeared for recreational purposes in an expository journal. Second, a mathematician's judgment on the permissibility of a gap presumably depends on whether the intended audience would be capable of bridging that gap. Hence, the target audience is a contextual feature of a proof that could influence a decision about the permissibility of a gap.

The third source arises from mathematicians disagreeing about which *inferential schemes are permissible* in a proof. For instance, one mathematician may accept a computer-assisted argument (i.e., an argument where some assertions are justified solely on behalf of computer calculations) as a proof on the grounds that it is permissible to warrant a claim by a computer calculation (e.g., Fallis, 1996, who claims that most mathematicians would accept some computer-assisted arguments as proofs). Another mathematician may reject the argument, not because she has doubts about the quality or correctness of how the argument was carried out, but because *in principle* arguments that rely on a computer to perform calculations cannot qualify as proofs (e.g., Rota, 1997). This is the type of disagreement that is of theoretical interest for our paper. The consensus view on proof asserts that mathematicians agree on which inferential schemes are permissible in a proof; the pluralistic view holds that mathematicians disagree on which inferential schemes are permissible.

2.4 The consensus view of proof

Some philosophers claim that mathematicians can usually agree on whether a particular mathematical argument is a proof (e.g., Berry, 2018). For instance, Azzouni (2004) offered an explanation of how "mathematicians are so good at agreeing with one another on whether some proof convincingly establishes a theorem" (p. 84). Mathematics educators have also endorsed this position. McKnight, Magid, Murphy, and McKnight (2000) asserted that "all agree that something is either a proof or it is not and what makes it a proof is that every assertion in it is correct" (p. 1). Selden and Selden (2003) remarked on "the unusual degree of agreement about the correctness of arguments and the truth of theorems arising from the validation process" (p. 7). Selden and Selden further suggested that mathematicians are consistent in judging the validity of a proof: "Mathematicians say an argument proves a theorem, not that it proves it for Smith and possibly not for Jones" (p. 11).

Philosophers who advance the consensus view on proof often draw strong conclusions about mathematical practice. For instance, many philosophers have argued that mathematics is a special discipline "with a type of knowledge being categorically more secure than that of other sciences" (Geist, Loewe, and van Kerkhove, 2010, p. 155). Berry (2018) claimed that established mathematical results are rarely subsequently overturned in part because there is consensus about which arguments are proven (although for a critical analysis, see Geist et al., 2010).

The consensus view on proof significantly shapes mathematics education research, affecting both theory and methodology. A central tenet of Harel and Sowder's (2007) influential proof schemes framework is "the goal of instruction must be unambiguous—namely, to gradually refine current students' proof schemes toward the proof scheme shared and practiced by contemporary mathematicians. *This claim is based on the premise that a shared proof scheme exists*" (p. 809; emphasis ours).

The consensus view is also a root presumption for mathematics education studies of alignment between students' views of proving and those held by mathematicians. For instance, one popular paradigm is to ask students to evaluate arguments and decide whether they qualify as proofs (e.g., Alcock & Weber, 2005; Healy & Hoyles, 2000; Ko & Knuth, 2013; Selden & Selden, 2003; Weber, 2010). If students judge a (purportedly) invalid argument to be a proof or if they judge a correct proof as invalid, the mathematics educators find this to be a mathematical shortcoming on the part of the student that warrants instructional remediation. These studies are sensible only if the mathematical community at large would agree with the mathematics educators' judgments on the validity of the arguments that they used. It would hardly be reasonable to critique a student for evaluating an argument as (not a) proof if mathematicians themselves could not reach a consensus on the item.

2.5 The pluralistic view of proof

Other philosophers are skeptical of the consensus view. In response to Azzouni (2004, cited above), Rav (2007) said that proof is "pluralistic" by nature and added, "because of the historical and methodological wealth of proof practices (plural), any attempt to encapsulate such multifarious practices in a unique and uniform one-block perspective is bound to be defective" (p. 299, the parenthetical remark was the author's). Aberdein (2009) coined the term proof* to denote the "species of alleged 'proof' where there is no consensus that the method provides proof, or there is a broad consensus that it doesn't, but a vocal minority or an historical precedent point the other way" (p. 1). As examples of proof*, Aberdein included "picture proofs*, probabilistic proofs*, [and] computer-assisted proofs*" (p. 1). Each of the proofs* cited above is based on a different inferential scheme that are purportedly permitted by some mathematicians but not others. For instance, the legitimacy of computer-assisted proofs* hinges on whether it is permissible to warrant a claim in a proof solely by a computer calculation. Rav's and Aberdein's commentaries question whether the mathematical community has reached consensus on the permissibility of every inferential scheme.

Many mathematics educators support the pluralistic position as well (e.g., Almeida, 1996; Dreyfus, 2004; Inglis et al., 2013). For instance, Dreyfus (2004) asserted that "what counts as a proof is not absolute […] It may thus depend on contextual factors such as historical period, the domain within mathematics and the aim of the proof in a given situation" (p. 4). To support his position that proof is not absolute, Dreyfus (2004) cited computer-assisted proofs and visual proofs among the specific types of proofs whose status as proofs among mathematicians is uncertain.

Advocates of the pluralistic view of proof sometimes critically question the pedagogical goals with regard to proof. First, Dreyfus (2004) challenged whether a goal of instruction for students should be developing specific requirements for what constitutes a proof. He argued that, "In view of the socially constituted nature of proof, even in mathematics, *uniformity is the last thing to aim for*" (p. 8, emphasis ours; note the contrast to Harel & Sowder, 2007). Second, a pluralistic view advises teachers to be aware that different judgments on particular proof candidates are possible and they should consider admitting arguments that use a broad range of inferential schemes as proofs (Almeida, 1996; Boero et al., 2018; Dreyfus, 2004).

The pluralistic view also has implications for conducting mathematics education research. Consider the proof evaluation studies in which students are asked to determine if various arguments qualify as proofs. Czocher and Weber (in press) and Inglis et al. (2013) claimed that if mathematicians do not agree on the status of proofs, items in proof evaluation studies would need to be tested with mathematicians before using them in studies of students' judgments about proof.

2.6 Motivation for study

The extent to which mathematicians agree on which inferential schemes are permissible within a proof has significant implications, both for how scholars understand the nature of mathematical practice and how mathematics should be taught. Yet, as we documented above, scholars have been unable to reach a consensus on this issue. In this paper, we present an exploratory study addressing the following questions: (1) To what extent do mathematicians agree on the validity of computer-assisted proofs and visual proofs in elementary number theory? (2) To what extent do mathematicians agree on the validity of proofs using methods of inference standard in a university introductory proof course, such as direct proof and proof by cases, in elementary number theory? Note that question (1) is an open question, with the consensus view predicting agreement and the pluralistic view predicting a lack of agreement.

Broadly, there are two ways to investigate mathematicians' perceptions on the validity of inferential schemes. A qualitative approach might study a small number of mathematicians' beliefs about validity of given arguments in a task-based or open-ended interview. A qualitative study could provide valuable nuance to views of mathematicians' thinking. Of course, the drawback would be that the small sample size would raise questions about representativeness and therefore limit generalization to the larger community of mathematicians. For this reason, we believe that a qualitative approach would be inappropriate for examining whether, or the extent to which, there is consensus or disagreement within the larger community. A survey approach could ask a large number of mathematicians about their views on specific arguments and generate tentative conjectures about the larger mathematical community's treatment of proof.

Our study used a survey approach. Our rationale is that the extant literature already offers individual mathematicians' insights about particular inferential schemes or whether mathematicians usually agree on whether an argument is a proof. For instance, Selden and Selden (2003), both published mathematicians, have described their experiences on how mathematicians usually reach a consensus on whether an argument is a proof. Rav (1999, 2007), also a published mathematician, made the case that disagreement about the validity of inferential methods was common. We believe these insights can be complemented by measuring when, and the extent to which, disagreement about whether an argument is a proof occurs amongst mathematicians. Our results will provide evidence regarding if and when mathematicians disagree, but further research will be necessary to uncover *why* the agreement or lack thereof occurs. We treat potential follow up studies in the discussion.

3 Methods

3.1 An internet study

A key aim of this study was to estimate the rates of agreement and disagreement in mathematicians' evaluations of specific arguments. Because a large sample of mathematicians is needed to make such estimates, we chose to collect data through an online study in order to increase the sample size of mathematicians. This method of data collection to study mathematical practice was pioneered by Inglis and Mejia-Ramos (2009a, 2009b) and has since been used as a primary means to conduct quantitative studies about the mathematical community (e.g., Fukawa-Connelly & Cook, 2016; Inglis & Mejia-Ramos, 2009a, 2009b; Lew, 2016; Weber, 2013; Weber & Mejia-Ramos, in press). Papers using this methodology have been published in prestigious journals[5], including *Research in Mathematics Education*. Psychologists have documented the validity of these studies (e.g., Kranz & Dalal, 2000; Gosling et al., 2004).

3.2 Participants

A recruitment e-mail was sent to the mathematics department faculty secretary at 25 large research universities in Great Britain. The email requested they forward an invitation to the mathematics staff to participate in a survey on what proofs were valid. This e-mail contained a hyperlink to a Qualtrics[6] survey. Ninety-four mathematicians participated and completed the survey.

---

[5] These journals include *Educational Studies in Mathematics* (Mejia-Ramos & Weber, 2014), *Cognition and Instruction* (Lai, Weber, & Mejia-Ramos, 2012; Inglis & Mejia-Ramos, 2009a), and *Research in Mathematics Education* (Weber, 2013).

[6] Qualtrics is a software company that allows researchers to conduct on-line surveys.

3.3 Items

The survey asked participants to evaluate the validity of five arguments, which are presented in the Appendix of this paper. We summarize the arguments in Table 1 below:

Type	Claim	Source	Inferential Methods
Prototypical Proof (PP1)	The n^{th} prime p_n satisfies $p_n \leq 2^{2^{n-1}}$ for all $n \geq 1$	Undergrad Textbook (Jones & Jones, 1998)	Proof by strong induction, algebraic manipulation of inequalities
Prototypical Proof (PP2)	If n is a number of the form $4k+3$, then n is not perfect	*American Math. Monthly* (Holdener, 2002)	Proof by cases, modular arithmetic
Empirical Proof (EP)	If n is an odd integer, then n^2 is an odd integer	*MAA Blog on teaching*[7] (Weber, 2003)	Naïve empiricism
Visual Proof (VP)	If n is an odd integer, then n^2 is congruent to 1 (mod 8)	*Math Horizons* (Nelsen, 2008)	Visual proof, recognizing features present in a generic diagram
Computer-Assisted Proof (CAP)	$\pi = \sum_{k=0}^{\infty} \frac{(-1)^k}{4^k}\left(\frac{2}{4k+1} + \frac{2}{4k+2} + \frac{1}{4k+3}\right)$	*American Math. Monthly* (Adamchik & Wagon, 1997)	Computer-assisted calculation to simplify a complicated integral

Table 1. Summary of items used in this study

We chose these arguments with the following considerations in mind. First, to ensure that the participants could follow the arguments, we chose arguments in the domain of undergraduate number theory that were presented in expository journals and undergraduate textbooks. This ensured that specialist knowledge was not needed to interpret any of the arguments. Second, we focused on a conceptual domain that is taught to undergraduates, often in a transition-to-proof course, to increase the relevance of our findings for mathematics educators. If it is the case that, say, algebraists and topologists have different conceptions of validity in elementary number theory, this would suggest

[7] https://www.maa.org/programs/faculty-and-departments/curriculum-department-guidelines-recommendations/teaching-and-learning/research-sampler-8-students-difficulties-with-proof. We chose to reference an educational blog for this empirical argument as we could not find a traditional mathematical outlet (i.e., a journal or a textbook) that would sanction an empirical argument as a proof.

that undergraduates are receiving mixed messages on what a proof is. Third, we deliberately chose some arguments that used typical inferential schemes in mathematical practice. We chose two “prototypical proofs” (PP1, PP2), which we call prototypical because they are arguments using standard mathematical notation and inferential schemes common in number theory proofs (e.g., proof by induction, case-based reasoning, inference by calculation). We chose two arguments that we anticipated might engender disagreement, a visual proof (VP) and a Computer Assisted Proof (CAP). We anticipated disagreement because previous scholarship has found them controversial (i.e., mathematicians would have strong and conflicting thoughts about the validity of these arguments): Aberdein (2009) suggested that they be considered as proofs* and Dreyfus (2004) stated that these kinds of proofs are controversial. Finally, we chose an empirical proof (EP), which we anticipated mathematicians would largely agree would not qualify as a proof.

These choices aligned with our intent to test the hypothesis that mathematicians would agree on which inferential schemes are permissible. In choosing the controversial proofs, we aimed to maximize the possibility that any disagreement we observed would be due to the inferential scheme used in the proof and simultaneously minimize the possibility that the disagreement was due to performance error or gap size. In the next sub-section, we describe how we minimized disagreements due to performance error. To minimize potential disagreements over gap size, we chose the Visual Proof and the Computer Assisted Proof that contained only a small number of inferences. In the Visual Proof, the theorem was deduced directly from the diagram. In the Computer Assisted Proof, the proof contained two inferences – a simplification of a complicated integral in

Mathematica and a trivial calculation. Because the trivial calculation would presumably be non-problematic to any mathematician, we believe that any disagreement about Visual Proof or Computer Assisted Proof would be due to the inferential scheme employed.

We used arguments from expository journals and undergraduate textbooks for two reasons. First, it would be unreasonable to ask mathematicians to read longer proofs from specialist research journals. Second, using discipline-specific proofs would mean we would need either to ask some mathematicians to evaluate proofs outside of their area of expertise or we would need different sets of proofs for each discipline. The former would not be a valid experimental design. The latter would produce very small sample sizes and would yield uninterpretable results because there would be no way to compare across subdisciplines in the case we found consensus within each subdiscipline. Nonetheless, these decisions may limit the generality of our study since mathematicians may referee proofs submitted to a research journal differently than how they would evaluate proofs from expository journals or from undergraduate textbooks. For instance, it is possible that mathematicians may place less value in rigor and more value in explanation when evaluating proofs from these sources. We discuss these possibilities in the Discussion section of this paper.

3.4 Procedure

At the start of the survey, participants were told that they would be asked to make validity judgments on five mathematical arguments from number theory. We took deliberate steps to ensure that participants focused on the inferential schemes in the argument and to minimize the possibility that disagreements were due to performance errors. To direct participants' attention to the inferential schemes that were used, the participants were told that the focus of the study was on the type of reasoning within the argument and that no attempt was being made to deceive them. They were then told that each proof was published, each sentence in the argument was true, and each calculation was carried out correctly. After each proof was presented, participants were asked to make four judgments:

- Participants were asked, "On a scale of 1 through 10, how representative is this argument to the types of proofs that you encounter in your mathematical practice? (10 signifies very typical. 1 signifies this proof bears little similarity to the proofs that you read and write)". Participants were given an open response box to reply. For each proof, a small number of participants (less than seven) provided non-numerical responses which we did not include in our analysis.
- Participants were asked, "If you were forced to choose, would you say that this argument is a valid proof? (As a reminder, you may assume the statements within the argument are true. We are interested in whether you think the type of reasoning to deduce the statements and the conclusion in the argument is valid)". Participants were given the following two options: (i) This is a valid proof; and (ii) This is not a valid proof. One purpose of asking this binary question is to show how participants' apparent disagreement on the validity of some mathematical arguments can be explained by their responses to the two subsequent questions.

- Participants were asked, “What percentage of mathematicians do you think would agree with your judgment above?” Participants were given the following four options: (i) 91-100%; (ii) 71-90%; (iii) 51-70%; (iv) 31-50%; and (v) 0-30%.
- Participants were asked, “For a more nuanced view of validity, which do you think best captures the validity of this argument?”. Participants were given the following four options: (i) I would consider this to be a valid proof in nearly all mathematical contexts.; (ii) I would generally consider this argument to be a valid proof, but there are some contexts where I would consider this argument to be invalid.; (iii) I would generally consider this argument to be invalid, but there are mathematical contexts in which I would consider this argument to be a valid proof.; (iv) I would consider this argument to be invalid in nearly all mathematical contexts.

4 Results

Among the participants completing the survey, 41 were regular faculty (22 with more than six years experience, 19 with less than six years), 10 were post-doctoral faculty, 41 were doctoral students, and the remaining two participants did not specify their status. Prior to analysis, statistical comparisons (*t*-tests for the representative ratings and Fisher exact tests for the other judgments) were made on all the ratings between the 41 doctoral students and the 51 post-docs and faculty members on each of their ratings for the five proofs with an alpha of .05. None of these statistical tests found a difference between the doctoral students and the other participants. Similarly, statistical comparisons determined that there was no statistically significant relationship between faculty status and their responses to the items with an alpha of .05. These results are consistent with the broader literature claiming that mathematicians' practice or epistemic stances are not significantly related to their experience or faculty status; in particular, in prior studies, doctoral students responded similarly to questions about mathematical practice as mathematics faculty (Inglis et al., 2013; Mejia-Ramos & Weber, 2014; Weber & Mejia-Ramos, in press). Consequently, we analyzed all data in aggregate.

The main results from this study are presented in Table 2 and Table 3 below.

Proof	Mean Typicality Rating	Validity Judgment		Anticipated	Level of	Agreement	
		Valid Proof	Invalid Proof	91-100%	71-90%	51-70%	0-50%
PP1	7.4	99%	1%	90%	9%	1%	0%
PP2	6.8	98%	2%	78%	20%	2%	0%
VP	2.6	62%	38%	14%	46%	33%	7%
CAP	2.7	39%	61%	10%	41%	37%	12%
EP	1.6	0%	100%	93%	0%	1%	6%

Table 2. Participants' judgment on the validity of the five proofs that they read

Proof	Valid proof in nearly all contexts	Valid proof but invalid in some contexts	Invalid proof but valid in some contexts	Invalid proof in nearly all contexts
PP1	94%	5%	1%	0%
PP2	79%	20%	0%	1%
VP	21%	33%	39%	6%
CAP	10%	33%	42%	15%
EP	1%	1%	3%	95%

Table 3. Participants' judgment on the more fine-grained view of validity

Prototypical proofs. The typicality ratings of these proofs were high (7.4, SD=2.5, for PP1 and 6.8, SD=2.6, for PP2), indicating that participants viewed these proofs as fairly typical. For PP1, there was a high level of agreement among the participants. All but one participant judged the proof to be valid. Further, most participants (94%) felt that the proof would be valid in nearly all mathematical contexts and most (90%) thought that more than 90% of their colleagues would agree with them. A similar (although less pronounced) trend was observed with PP2. All but two participants judged the proof to be valid, most (79%) thought the proof was valid in nearly all contexts, and most (78%) thought that more than 90% of their colleagues would agree with their judgment.

Empirical proof. The level of agreement for the Empirical Proof was also high. Every participant evaluated the Empirical Proof to be invalid; most indicated that this argument would be invalid in nearly all mathematical contexts (95%) and that over 90% of their colleagues would agree with their evaluation (92%).

Non-prototypical proofs. The participants did not view the Visual Proof and the Computer Assisted Proof as typical proofs, giving typicality ratings of 2.6 (SD=1.8) and 2.7 (SD=2.1) respectively. There was dissent among the participants as to whether the Visual Proof and the Computer Assisted Proof were valid; 62% judged the Visual PRoof to be valid and 39% judged the Computer Assisted Proof to be valid. However, most participants were aware that there would be dissent among mathematicians as to the validity of these proofs. In each case, fewer than 15% of the participants predicted that over 90% of their colleagues would agree with their judgment. Particularly important was participants' judgment about the validity of the proof depending on context. In both cases, most participants claimed that there were both situations in which the proof would be valid and situations where the proof would be invalid (72% for the Visual Proof, 75% for the Computer Assisted Proof, these numbers are obtained by adding the middle two columns of Table 2). If we re-framed the question about the Visual Proof and the Computer Assisted Proof to be, "are there contexts in which this argument would be a valid proof?", we would anticipate a high level of agreement. We can predict participants' response to this question by adding the first three columns of Table 2 (100% minus the value in the "invalid in most contexts" category) from which we obtain values of 85% for the Computer Assisted Proof and 94% for the Visual Proof. Consequently, the subsequent questions suggest that the variation participants exhibited in their judgments of the validity of the Computer Assisted Proof and the Visual Proof may be due to the binary judgment that they were required to make.

One post-hoc hypothesis that we considered was that participants' judgments on typicality and validity were synonymous—that is, participants simply judged all typical proof methods to be valid and atypical proof methods to be invalid. There was a relationship between participants' typicality ratings for the Computer Assisted Proof and the Visual Proof and whether they thought the proofs were valid. Of the 15 participants who gave the Computer Assisted Proof a typicality rating of 5 or higher, 11 (73%) evaluated the Computer Assisted Proof as valid. Of the 74 participants who gave the Computer Assisted Proof a typicality rating of 4 or lower, 25 (34%) evaluated the Computer Assisted Proof as valid, a significant difference (Fisher exact test $p = .0427$). (Five participants gave a non-numerical response for typicality). Similarly, of the 14 participants who gave the Visual Proof a typicality rating of 5 or higher, 12 (86%) evaluated the Visual Proof as valid. Of the 78 participants who gave the Computer Assisted Proof a typicality rating of 4 or lower, 44 (56%) evaluated the Visual Proof as valid, a significant difference (Fisher exact test $p = .0079$). (Two participants gave a non-numerical response for typicality). Nonetheless, the data suggest that even for participants who viewed Computer Assisted Proof and the Visual Proof as typical, these participants were still more likely to judge these proofs as invalid than PP1 and PP2.

5 Discussion

5.1 Summary of main findings

We summarize our main findings and how they relate to the consensus view and pluralistic view on proof. Our main finding is that mathematicians disagreed on the status of computer-assisted proofs and visual proofs, which is consistent with the pluralistic position and confirms Aberdein's (2009) and Dreyfus' (2004) statements about mathematical practice. This is significant for two reasons. First, some scholars have claimed that computer-assisted proofs and visual proofs are *not* controversial. For instance, Fallis (1996) wrote that a debate about the status of computer-assisted proofs was "archaic" on the grounds that such proofs were usually accepted. Inglis and Mejia-Ramos (2009b) asserted that the "common view" was that visual inferences could not be used to reliably secure mathematical knowledge and consequently could not be admissible in a proof. In the latter case, the fact that 62% of our participants accepted the Visual Proof as a proof and 94% thought the Visual Proof would be valid in at least some contexts demonstrates that the common view might not be so common after all. (We found similar findings when mathematicians were asked to judge the validity of graphical inferences in a student-generated proof in a real analysis course; see Weber and Mejia-Ramos, in press). Second, the disagreement that we observed poses a challenge for mathematics educators who ascribe to a consensus view. A consensus view is a critical assumption underlying the proof schemes framework (Harel & Sowder, 2007). That is, finding support for the pluralist view calls into question the use of proof schemes as a basis for instructional theory. Happily, our next main result provides a response to this challenge.

Our next main result is that the disagreement that we observed only occurred for inferential methods that mathematicians found to be atypical. Further, mathematicians were aware that these proofs would be controversial. There was near universal agreement that PP1 and PP2 were valid proofs and the Empirical Proof was not. We believe this too is significant. First, it suggests a way in which the consensus view is consistent with many educators' and philosophers' observations: mathematicians may indeed agree to a remarkable degree on the proofs that they *typically* encounter. Hence, claims such as "the unusual degree of agreement about the correctness of arguments and the truth of theorems arising from the validation process" (Selden & Selden, 2003, p. 7) are plausible provided that this is interpreted as an assertion about the typical proofs that mathematicians encounter. Second, the types of reasoning we aim to teach students to use (i.e., proof by induction, proof by cases, algebraic manipulation, modular reasoning) are types of reasoning that were judged as valid by nearly all mathematicians. Likewise, the mathematicians uniformly rejected naïve empirical reasoning which students are generally instructed not to use. Thus, even though mathematicians disagree on the status of *some* types of inferences, there appears to be a consensus on the types of inferences that form the foundation of instruction. In short, Harel and Sowder's (2007) assumption about mathematicians' shared proof schemes is supported empirically in the context of the inferences that we expect students to use or avoid, which is a wide enough scope for their theory to productively inform pedagogy. Third, many scholars have noted that the remarkable level of agreement among mathematicians about proof validity is unusual; in most other disciplines, it is rare that a single argument can command universal assent and settle an open question. Even if this does not *always* happen in mathematics (i.e., there are many high profile examples of conjectures whose status remained murky after a purported proof was produced), this does occur regularly. Hence, we agree with Azzouni (2004) and Berry (2018) that this is a phenomenon worthy of further investigation.

5.2 Limitations and caveats

Our study had two significant limitations, each of which suggests useful avenues for future research. First, while our data suggests that mathematicians *do* agree on the validity of inferential schemes in prototypical proofs and disagree on the validity of inferential schemes that are not typical, the data does not help us understand *why* this is the case. A follow-up interview study in which mathematicians are shown the results of this paper and asked to provide their interpretation of these results could provide insight into why we obtained the results that we did. For instance, perhaps particular groups of mathematicians, such as topologists or geometers, are more accepting of visual proofs than others, which would be consistent with the Computer Assisted Proof and the Visual Proof data presented at the end of the results section. Addressing this issue is beyond the scope of our methodology (i.e., it would be extremely difficult to get sufficiently large samples of mathematicians from various sub-disciplines to make comparisons of this type), but we believe that qualitative studies can complement this quantitative study and provide depth and nuance into the findings we reported.

Second, our conclusions are based on an implicit assumption: because the mathematicians in our sample agreed on the validity of the two prototypical proofs in our study, they would reach a similar level of agreement on other prototypical proofs that they encountered. However, only two prototypical proofs were used in this study. Further, these mathematicians' evaluations were made in the context of proofs that appeared in expository journals or undergraduate textbooks. It is possible that the disagreements with the Visual Proof and the Computer Assisted Proof were due to differing views of what was appropriate for a general audience or even who members of that general audience would be; however, such disagreements would not occur in a specialist journal that was written for active researchers.

This hypothesis about the role of context can be tested in a future study in the following way. Mathematicians could be asked to bring in a recent issue of a mathematical journal that they regularly read. For each proof that appeared in the journal, the mathematician could be asked if the methods used in the argument were typical and given that the methods in the argument were carried out correctly, would the argument be a valid proof to the mathematician. Would most other mathematicians, including those from other disciplines, also accept the proof as valid if they could follow the proof and were convinced that the proof contained no mistakes? Our prediction is that participants would say that most published proofs would use typical inferential schemes, that participants would usually evaluate them as valid (modulo correct implementation of the methods), and participants would believe that their peers would make the same judgment.

5.3 Implications for mathematics education

Recall that Harel and Sowder's (2007) unambiguous goal for instruction was for students' proof schemes to align with the proof schemes of contemporary mathematicians; this goal presumes that contemporary mathematicians share a common view on acceptable proof schemes. Recall also that Dreyfus (2004) objected: because mathematicians disagree about what constitutes a proof, uniformity should not be a goal of instruction. What our work suggests is that there are three categories of inferential schemes: (i) valid schemes that are permissible in a proof such as proof by induction; (ii) invalid schemes that are never permissible in a proof such as naïve empiricism (Balacheff, 1988); and (iii) controversial schemes whose permissibility is unclear, such as drawing inferences from diagrams. Our opinion is that Harel and Sowder's (2007) aims of having students see the limitations of invalid schemes and the legitimacy of valid schemes (e.g., Harel, 2001) is appropriate. Uniformity for these schemes is a worthwhile goal for instruction because mathematicians have reached consensus on the status of these inferential schemes. For controversial schemes, we accept Dreyfus' (2004) critique; we believe students should be aware that the permissibility of these inferential schemes is contextual and subjective (c.f., Boero et al., 2018). As we argue in Czocher and Weber (in press), instead of debating whether arguments containing these inferences are proof, a more productive discussion (in classrooms and among mathematics educators) should concern the affordances and limitations that these types of inferences provide.

Acknowledgments. We are grateful to Matthew Inglis, Kristen Lew, Kate Melhuish, and the anonymous reviewers for helpful comments on earlier drafts of this manuscript.

6 References

Aberdein, A. (2009). Mathematics and argumentation. *Foundations of Science*, 14, 1-8.

Adamchik, V. & Wagon, S. (1997). A simple formula for π. *American Mathematical Monthly*, 104, 852-855.

Alcock, L. & Weber, K. (2005). Proof validation in real analysis: Inferring and validating warrants. *Journal of Mathematical Behavior*, 24(2), 125-134.

Almeida, D. (1996). Variation in proof standards: Implications for mathematics education. *International Journal of Mathematical Education in Science and Technology*, *27*(5), 659-665.

Azzouni, J. (2004). The derivation-indicator view of mathematical practice. *Philosophia Mathematica*, *12*, 81-106.

Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 216–235). London, United Kingdom: Hodder & Stoughton.

Balacheff, N. (1999). For an ethnomathematical questioning on the teaching of proof. *International Newsletter on the Teaching and Learning of Mathematical Proof.* Downloaded from: http://www.lettredelapreuve.org/OldPreuve/Newsletter/990910Theme/990910ThemeUK.html

Balacheff, N. (2008). The role of the researcher's epistemology in mathematics education: an essay on the case of proof. *ZDM*, *40*(3), 501-512.

Berry, D. (2018). Proof and the virtues of shared enquiry. *Philosophia Mathematica*, 26, 40-58.

Boero, P. (1999). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on the Teaching and Learning of Mathematical Proof.* Downloaded from: http://www.lettredelapreuve.org/OldPreuve/Newsletter/990708Theme/990708ThemeUK.html

Boero, P., Fenaroli, G., & Guala, E. (2018). Mathematical argumentation in elementary teacher education: The key role of the cultural analysis of the content. In A. Stylianides & G. Harel (Eds.) *Advances in mathematics education research on proof and proving*. (pp. 49-67). Springer: Dordrecht.

CadwalladerOlsker, T. (2011). What do we mean by mathematical proof? *Journal of Humanistic Mathematics*, 1, 33-60.

Cirillo,M., Kosko, K., Newton, J., Staples, M., & Weber, K. (2015). Conceptions and consequences of what we call argumentation, argument, and proof. In T. Bortell & K. Bieda (Eds.) *Proceedings of the 37th annual meeting of the North American chapter of the psychology of mathematics education* (pp. 1343-1351).

Cook, J. P., & Fukawa-Connelly, T. (2015). The Pedagogical Examples of Groups and Rings That Algebraists Think Are Most Important in an Introductory Course. *Canadian Journal of Science, Mathematics and Technology Education*, *15*(2), 171-185.

Czocher, J. & Weber, K. (in press). Proof as a cluster concept. To appear in *Journal for Research in Mathematics Education.*

Dawkins, P. & Weber, K. (2017). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics*, 95, 123-142.

Devlin, K. (2003). The shame of it all. *Devlin's Angle*. Retrieved from http://www.maa.org/external_archive/devlin/devlin_06_03.html

Dreyfus, T. (2004) What counts as proof in the mathematics classroom? In Kourkoulos, M., Troulis, G. & Tzanakis, C. (Eds.), *Proceedings of the 3rd Colloquium on the Didactics of Mathematics* (pp. 114-132). Rethymnon, Crete: Department of Education, University of Crete.

Duval, R. (1999). Questioning argumentation. *International Newsletter on the Teaching and Learning of Mathematical Proof.* Downloaded from: http://www.lettredelapreuve.org/OldPreuve/Newsletter/991112Theme/991112ThemeUK.html

Duval, R. (2007). Cognitive functioning and the understanding of mathematical processes of proof. In P. Boero (Ed.), *Theorems in school*: *From historic, epistemology, and cognition to classroom practice* (pp. 163–181). Rotterdam: The Netherlands: Sense.

Fallis, D. (1996). Mathematical proof and the reliability of DNA evidence. *American Mathematical Monthly*, 103, 491-497.

Geist, C., Löwe, B., & Van Kerkhove, B. (2010). Peer review and testimony in mathematics. In B. Löwe & T. Müller (Eds.), *Philosophy of mathematics: Sociological aspects and mathematical practice* (pp. 155-178). London: College Publications.

Gosling, S.D., Vazire, S., Srivastava, S., & John, O.P. (2004). Should we trust web-based studies? A comparative analysis of six studies about internet questionnaires. *American Psychologist*, *59*, 93-104.

Harel, G. (2001). The development of mathematical induction as a proof scheme: A model for DNR-based instruction. In S. Campbell & R. Zaskis (Eds.), *Learning and teaching number theory*: *Research in cognition and instruction* (pp. 185–212). Dordrecht, The Netherlands: Kluwer.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.) *Second handbook of research on mathematics teaching and learning,* (pp. 805-842). Reston, VA: NCTM.

Healy, L. & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*(4), 396-428.

Herbst, P. & Balacheff, N. (2009). Proving and knowing in public: The nature of proof in a classroom. In M. Blanton, D. Stylianou, & E. Knuth (Eds.) Teaching and learning proof across the grades: A K-16 perspective (pp. 40-64). New York: Routledge.

Holdener, J. (2002). A theorem of Touchard on the form of odd perfect numbers. *American Mathematical Monthly*, 109, 661-663.

Inglis, M. & Aberdein, A. (2016). Diversity in proof appraisals. In B. Larvor (Ed.) *Mathematics Cultures*. (pp. 163-179). Dordrecht: Springer.

Inglis, M. & Mejia-Ramos, P. (2009a). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction*, 27, 25-50.

Inglis, M. & Mejia-Ramos, P. (2009b). On the persuasiveness of visual arguments in mathematics. *Foundations of Science,* 14, 97-110.

Inglis, M., Mejia-Ramos, J.P. Weber, K. & Alcock, L. (2013). On mathematicians' different standards when evaluating elementary proofs. *Topics in Cognitive Science*, 5, 270-282.

Jones, G. & Jones, M. (1998). *Elementary number theory*. Springer: London

Ko, Y. Y., & Knuth, E. J. (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *The Journal of Mathematical Behavior*, *32*(1), 20-35

Kranz, J. and Dalal, R. (2000). Validity of web-based psychological research. In M. Birnbaum (Ed.) *Psychological experiments on the internet*. (pp. 35-60). Academic Press: San Diego, CA.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery.* Cambridge: Cambridge University Press.

Lai, Y., Weber, K. & Mejia-Ramos, J.P. (2012). Mathematicians' perspectives on features of a good pedagogical proof. *Cognition and Instruction*, 30, 146-169.

Lew, K. (2016). Conventions of the language of mathematical proof writing at the undergraduate level (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick).

Maher, C. A., & Martino, A. M. (1996). The development of the idea of mathematical proof: A 5-year case study. *Journal for Research in Mathematics Education, 27,* 194-214.

Mariotti, M. A. (2006). Proof and proving in mathematics education. In A. Gutierrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 173-204). Rotterdam, The Netherlands: Sense Publishers.

Mejia-Ramos, J.P. & Weber, K. (2014). How and why mathematicians read proofs: Further evidence from a survey study. *Educational Studies in Mathematics*, 85, 161-173.

McKnight, C., Magid, M., Murphy, T.J., & McKnight, M. (2000). *Mathematics education research: A guide for the research mathematician*. American Mathematical Society: Washington, D.C.

Nelsen, R. (2008). Visual gems in number theory. *Math Horizons*, 31, 7-9.

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics*, *66*, 23-41.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, *7*(1), 5-41.

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

Reid, D. & Knipping, C. (2010). *Proof in mathematics education*. Sense Publishers: Rotterdam.

Rota, G. C. (1997). The phenomenology of mathematical beauty. *Synthese*, *111*(2), 171-182.

Selden, A. & Selden, J. (2003). Validations of proofs written as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, 36(1), 4-36.

Staples, M. E., Bartlo, J., & Thanheiser, E. (2012). Justification as a teaching and learning practice: Its (potential) multifacted role in middle grades mathematics classrooms. *The Journal of Mathematical Behavior*, *31*, 447-462.

Steiner, M. (1978). Mathematical explanations. *Philosophical Studies, 34*(2), 135–151.

Stylianides, A. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, 38, 289-321.

Stylianides, A., Bieda, K., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In In A. Gutierez, G. Leder, & P. Boero (eds.) *2nd Handbook on the Psychology of Mathematics Education*. (pp. 315-351). Rotterdam: Sense Publishers.

Stylianides, G., Stylianides, A. & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (ed.) *Compendium for Research in Mathematics Education*. Reston, VA: National Council of Teachers of Mathematics.

Weber, K. (2003). Students' difficulty with proof. In A. Selden and J. Selden (eds.) *Research Sampler,* 8. Available from the following MAA website on the teaching of learning of mathematics: http://www.maa.org/t_and_l/sampler/research_sampler.html, Last downloaded: January 14, 2019.

Weber, K. (2010). Mathematics majors’ perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, 12, 306-336.

Weber, K. (2013). On the sophistication of naïve empirical reasoning: Factors influencing mathematicians’ persuasion ratings of empirical arguments. *Research in Mathematics Education*, 15, 100-114.

Weber, K. & Dawkins, P. (2018). Toward an evolving theory of Mathematical Practice informing Pedagogy: What standards for this research paradigm should we adopt? In A. Stylianides & G. Harel (Eds.) *Advances in mathematics education research on proof and proving*. (pp. 69-82). Dordrecht: Springer.

Weber, K., Inglis, M. & Mejia-Ramos, J.P. (2014). How mathematicians obtain conviction: Implications for mathematics instruction and epistemic cognition. *Educational Psychologist,* 49, 36-58.

Weber, K. & Mejia-Ramos, J.P. (in press). An empirical study on the admissibility of graphical inferences in mathematical proofs. To appear in A. Aberdein & M. Inglis (Ed.) *Advances in Experimental Philosophy of Logic and Mathematics*. London: Bloomsbury.

Appendix- Proofs used in this study

Prototypical Proof 1 (PP1)

Theorem: The nth prime p_n satisfies $p_n \leq 2^{2^{n-1}}$ for all $n \geq 1$.

Proof:

To show that the theorem is true, use strong induction on n. The result is true for n=1, since $p_1 = 2 \leq 2^{2^0} = 2$. Now assume the result is true for n = 1, 2, ..., k. Consider $N = p_1 p_2 \dots p_k + 1$. N must be divisible by some prime p and this prime cannot be $p_1, p_2, \dots p_k$ for then it would divide 1, which is impossible. Now, this new prime must be at least as large as the $(k+1)^{\text{th}}$ prime p_{k+1}, so

$$\begin{aligned} p_{k+1} &\leq p \\ &\leq p_1 p_2 \dots p_k + 1 \\ &\leq 2^{2^0} 2^{2^1} \dots 2^{2^{k-1}} + 1 \\ &= 2^{1+2+4+\cdots+2^{k-1}} + 1 \\ &= 2^{2^k - 1} + 1 \\ &= \frac{1}{2} 2^{2^k} + 1 \\ &\leq 2^{2^k}. \end{aligned}$$

Therefore the result is true for n=k+1. So the result is true for all $n \geq 1$.

Adapted from G. Jones & M. Jones (1998) *Elementary Number Theory*. Published by Springer.

Prototypical Proof 2 (PP2).

Theorem: If n is a number of the form $4k$+3, then n is a not perfect number.

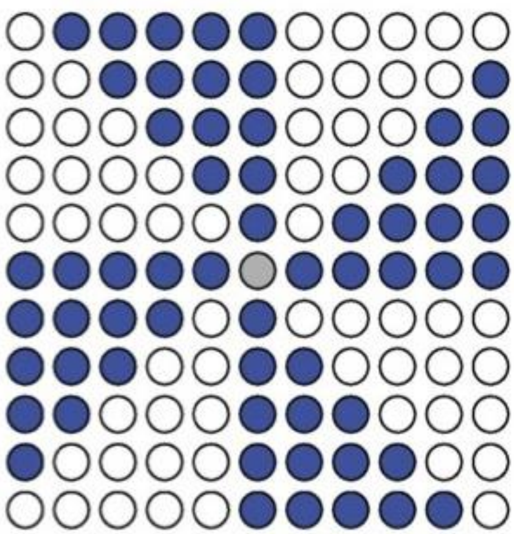
Proof: Assume n is a positive integer of the form $4k$+3. Then n is not a square. If d and $\left(\frac{n}{d}\right)$ are divisors of n, then either d<$\sqrt{n}$ or $\left(\frac{n}{d}\right) < \sqrt{n}$. Without loss of generality, assume d<$\sqrt{n}$. Since $n = d\left(\frac{n}{d}\right) \equiv 3(\text{mod } 4)$, either $d \equiv 3(\text{mod } 4)$ and $(\frac{n}{d}) \equiv 1(\text{mod } 4)$ or $d \equiv 1(\text{mod } 4)$ and $\left(\frac{n}{d}\right) \equiv 3(\text{mod } 4)$. Either way, $d + (\frac{n}{d}) \equiv 0(\text{mod } 4)$ and $\sigma(n) = \sum_{d|n, d<\sqrt{n}} d + \frac{n}{d} \equiv 0(\text{mod } 4)$. Computing $2n$=2($4k$+3) $\equiv 2(\text{mod } 4)$, we see that n cannot be perfect.

Published in: J. Holdener (2002). A theorem of Touchard on the form of odd perfect numbers. *American Mathematical Monthly*, 109, 661-663.

(Note: $\sigma(n)$ is the sum of the divisors of n. n is a perfect number if and only if $\sigma(n) = 2n$)

Visual Proof

Theorem: If n is an odd natural number, then $n^2 \equiv 1(\text{mod } 8)$
Proof: The proof is given in the picture below.



From R. Nelsen (2008). Visual gems in number theory. *Math Horizons*, 31, 7-9.

Computer-Assisted Proof the Computer Assisted Proof

Theorem: $\sum_{k=0}^{\infty}\left(\frac{-1}{4}\right)^k\left(\frac{2}{4k+1}+\frac{2}{4k+2}+\frac{1}{4k+3}\right)=\pi$

Proof: In this proof, we use *Mathematica* to assist with our computations.
Using *Mathematica,* we can verify that

$\sum_{k=0}^{\infty}\left(\frac{-1}{4}\right)^k\left(\frac{a_1}{4k+1}+\frac{a_2}{4k+2}+\frac{a_3}{4k+3}\right)=$

$\frac{a_2}{2}\pi+\left(\frac{a_1}{2}-a_2+a_3\right)\arctan(2)+\left(\frac{a_1}{4}-\frac{a_3}{2}\right)\ln(5).$

Setting a_1=2, a_2=2, and a_3=1 yields $\sum_{k=0}^{\infty}\left(\frac{-1}{4}\right)^k\left(\frac{2}{4k+1}+\frac{2}{4k+2}+\frac{1}{4k+3}\right)=\pi$, completing the proof.

Adapted from V. Adamchik and S. Wagon (1997). A simple formula for π. *American Mathematical Monthly*, 104, 852-855.

(Note: The authors noted that *Mathematica* could provide a verification for the correctness of the computations above).

Empirical Proof

An undergraduate in a transition-to-proof course produced the following proof:
Theorem: If n is an odd natural number, then n^2 is odd.
Proof: $1^2 = 1$, which is odd. $3^2 = 9$, which is odd. $5^2 = 25$, which is odd. I am convinced that this trend will continue. Therefore, if n is odd, then n^2 is odd.

Adapted as an example of a student proof from K. Weber (2003) *MAA Research Sampler on Undergraduate Mathematics Education*, published on the MAA's website.