

Conviction, validity, and proof

Mathematics majors' perceptions of conviction, validity, and proof

Keith Weber

Rutgers University

Address: 10 Seminary Place

Graduate School of Education

Rutgers University

New Brunswick, NJ 08901

U.S.A.

E-mail: [keith.weber@gse.rutgers.edu](mailto:keith.weber@gse.rutgers.edu)

Phone number: 732 932 7496

## Conviction, validity, and proof

### Mathematics majors' perceptions of conviction, validity, and proof

Abstract. In this paper, 28 mathematics majors who completed a transition-to-proof course were given ten mathematical arguments. For each argument, they were asked to judge how convincing they found the argument and whether they thought the argument constituted a mathematical proof. The key findings from this data were: (a) most participants did not find the empirical argument in the study to be convincing or to meet the standards of proof, (b) the majority of participants found a diagrammatic argument to be both convincing and a proof, (c) participants evaluated deductive arguments not by their form but by their content, but (d) participants often judged invalid deductive arguments to be convincing proofs because they did not recognize their logical flaws. These findings suggest improving undergraduates' comprehension of mathematical arguments does not depend on making undergraduates aware of the limitations of empirical arguments but instead on improving the ways in which they process the arguments that they read.

### 1. Introduction

In the last two decades, the role of proof has received increased emphasis within the mathematics education community. In reform-oriented mathematics, proof is expected to play a central role. In the United States, the National Council of Teachers of Mathematics (2000) argues that all students should “recognize reasoning and proof as fundamental to mathematics” and “develop and evaluate mathematical arguments and proofs” by the time they complete 12<sup>th</sup> grade (p. 56). Proof assumes an even greater role for undergraduate mathematics majors. In advanced mathematics courses, proof is a central means of conveying mathematical information and students’ ability to construct proof is a primary means of assessing their performance. Unfortunately, numerous studies illustrate that students at all levels have serious difficulties with many aspects of proof, including the construction and evaluation of proof and understanding the role that proof plays in mathematics (e.g., Bell, 1976; Chazan, 1993; Harel & Sowder, 1998, 2007; Healy & Hoyles, 2000; Senk, 1985; Weber, 2001).

In an influential article, Harel and Sowder (1998) coined the term *proof schemes* to refer to the ways in which a student attempts to convince herself or persuade others about the truth of a mathematical assertion. Many student difficulties with proof stem from the fact that students’ proof schemes are frequently different from the proof schemes held by contemporary mathematicians; in other words, students may be convinced by empirical or diagrammatic arguments and be uncertain about what features an argument must have to be acceptable to their teacher or the broader mathematical community. Harel and Sowder (2007) suggest that the notion of proof schemes can serve as a unifying framework for much of the diverse research on proof.

In the last two decades, there has been substantial research on what *types of arguments* students (and teachers) find convincing and what types of arguments they consider to be mathematical proofs. By types of arguments, I am referring to the types of evidence contained in the argument included to convince the reader about the veracity of the theorem being proved. This literature will be reviewed in detail in the next section, but in general researchers have concluded that students in middle school and high school, as well as first year university students and some mathematics teachers, are often convinced by empirical arguments<sup>1</sup>, believe the format of an argument is more important than its content in judging whether an argument is a proof, and do not appreciate the role that proof serves in mathematics (e.g., Martin & Harel, 1989; Chazan, 1993; Schoenfeld, 1989; Coe & Ruthven, 1994; Healy & Hoyles, 2000; Knuth, 2002; Knuth, Choppin, & Bieda, 2009).

There has been little systematic data collected on what types of arguments upper-level mathematics majors find convincing and what types of arguments they consider to be mathematical proof (see Hemmi, 2006, for a notable exception to what types of arguments students consider to proofs). In the United States, such students typically take a transition-to-proof course (see Moore, 1994; Alcock, 2009) in which they are taught basic ideas about proof. Exploratory studies indicate that students find this course to be very difficult and struggle to write proofs when the course is completed (e.g., Moore, 1994; Weber, Alcock, & Radu, 2007). However the proof schemes that students hold upon completion of this course have not been examined.

---

<sup>1</sup> Throughout this paper, the term *empirical argument* refers to an argument in support of a general assertion because the assertion was verified for one or several elements of the set. This differs from Harel and Sowder (1998), who refer to such arguments as being indicative of *inductive proof schemes*. To Harel and Sowder (1998), *empirical proof schemes* include both the tendency to gain conviction via examples and conviction by perceptual evidence.

The purpose of this paper is to address the following questions:

- What types of arguments do undergraduate mathematics majors who have completed a transition-to-proof course find convincing?
- What types of arguments do these same students consider to be a mathematical proof?
- Do these students ever have doubts that a proven claim is necessarily true? If so, why?

These questions are important for two reasons. First, advanced mathematics courses are typically taught in a “definition-theorem-proof” format (e.g., Dreyfus, 1991; Weber, 2004). Much of the lectures in advanced mathematics courses consist of professors presenting proofs of theorems to students. As Selden and Selden (1995, 2003) argue, if students are convinced by the wrong types of arguments or are unable to evaluate the correctness of a proof, then they cannot hope to gain legitimate conviction or understanding from the proofs that are presented to them in their upper-level mathematics classes. Consequently their experience attending their lectures and reading their textbooks will be considerably impoverished. Second, many mathematics majors are prospective high school teachers. The increased emphasis on proof in the elementary and secondary mathematics curriculum has placed a substantial burden on high school mathematics teachers (e.g., Knuth, 2002). In their *Principles and Standards for School Mathematics*, the NCTM (2000) advocates that “with guidance, students should develop high standards for accepting explanations” and that students “seek, formulate, and critique explanations so that classes become communities of inquiry” (p. 346). To help accomplish these goals, teachers should discuss the logical structure of the arguments that

## Conviction, validity, and proof

students present and assist students in critiquing each other's arguments (p. 346). Without a clear understanding of what constitutes a convincing and acceptable mathematical argument, teachers cannot be expected to accomplish these goals, or even properly evaluate the proofs that students submit on assessments (Weber, 2008). If prospective high school teachers complete a major in mathematics without a proper understanding of proof, they will be unable to accomplish the goals described above when they begin their teaching.

## 2. Related literature

### 2. 1. The types of arguments that students produce

Several studies have examined what types of arguments students produce when they are asked to justify a conjecture or prove why an assertion is true. Knuth, Choppin, and Bieda (2009) presented approximately 400 middle school children with six justification tasks in number theory and found that these tasks elicited empirical justifications 36% to 81% of the time, depending upon the task given to the students. Eighth grade students in Knuth, Coppin, and Beida's (2009) study continued to produce empirical arguments a substantial portion of the time despite using a reform-oriented curriculum whose goal was to have students produce deductive arguments by the end of middle school.

In a large-scale study, Healy and Hoyles (2000) examined the responses of nearly 2,500 secondary students in the UK when they were asked to prove both a familiar and an unfamiliar conjecture in high school algebra. They found that 34% of the responses for the familiar conjecture and 43% for the unfamiliar conjecture were empirical arguments.

## Conviction, validity, and proof

Healy and Hoyles also presented these students with two arguments purporting to demonstrate that other mathematical assertions were true and asked them to rate which argument was most similar to their own approach. For these two tasks, 24% and 39% of the students chose an empirical argument. Like the Knuth et al (2009) study, the students in Healy and Hoyles' (2000) study were using a reform-oriented curriculum that emphasized the importance of proof. Hoyles and Healy (2007) found similar results when students completed analogous tasks in geometry; their results are also consistent with a small-scale study of UK high school students conducted by Coe and Ruthven (1994).

At the university level, Recio and Godino (2001) asked over 400 first-year students to prove a statement from algebra and geometry and found that roughly 40% of the responses for these tasks were empirical. Although there have not been systematic studies of how mathematics majors perform on such tasks later in their undergraduate education, Harel and Sowder (1998) have interviewed students in advanced mathematics courses in a variety of settings and found that they too often attempt to prove by example. However, in several studies that my colleagues and I have conducted with upper-level mathematics majors (Weber, 2001; Weber & Alcock, 2004; Weber, Alcock, & Radu, 2007), we have found that empirical arguments are relatively rare responses when challenging proof construction tasks are presented to mathematics majors.

Most research suggests that students find empirical arguments convincing and believe they constitute an acceptable form of mathematical proof. However, the connection between the types of arguments that students construct and what they actually believe constitutes a mathematical proof is not straightforward. As Vinner (1997) warns, it is dangerous to infer information about students' cognitive structures by the incorrect

## Conviction, validity, and proof

answers that students produce, since the students may have provided these answers, not because they believed these answers were right, but for other social or pragmatic reasons, such as obtaining partial credit, pleasing the teacher or interviewer, or simply having no other idea about how to productively engage with their task. Hence it seems inappropriate to infer that students are convinced by empirical arguments simply because they posited one as a proof. Students may have resorted to doing so simply because they could not see how a deductive argument would proceed. Indeed, Knuth, Choppin, and Bieda (2009) found that middle school students were more likely to produce empirical arguments for complicated statements than for simpler statements. Their hypothesis is that for the complicated statements, “students likely had no recourse but to use examples as their means of justification (i.e., a verbal or symbolic argument may have been too difficult)” (p. 161). Healy and Hoyles (2000) also argue that, “empirical arguments predominated in students’ own proof constructions, although most students were aware of its limitations” (p. 396). In a recent study, Stylianides and Stylianides (2009) found a similar result, with pre-service teachers justifying a proposition with an empirical argument, even though many were aware that such a justification was not fully convincing.

Analogously, the absence of empirical arguments does not necessarily imply that students do not find these arguments convincing. The work of Segal (2000) suggests students often recognize that empirical arguments do not constitute proofs yet still find them convincing and Healy and Hoyles (2000) found that many high school students who personally prefer empirical arguments recognize that they would not receive high marks from their teachers. With this in mind, it is not clear whether the empirical arguments that students produce indicate that they find empirical arguments to be convincing, they find

## Conviction, validity, and proof

the arguments to be an acceptable form of proof, or that they lack the skills to form a deductive justification.

### 2. 2. What types of arguments do students find convincing?

As noted above, inferring students' proof schemes from the arguments they produce is problematic, making it challenging for researchers to determine what types of arguments students actually find convincing. The approach used in this paper was to present students with arguments in support of a proposition and ask them to determine how convincing they find these arguments. Several researchers (e.g., Selden & Selden, 2003; Hazzan & Zaskis, 2003; Mamona-Downs, 2005; Weber, 2008) have noted that although there have been many studies in which students were asked to construct proofs, the literature on how students read proofs has been comparatively sparse.

At the university level, Segal (2000) interviewed over 30 first-year university students studying mathematics in the UK and found that 70% of these students were convinced of a claim in linear algebra by an argument that verified that it held for a single example. Further, completing a year of university math classes did little to change these evaluations. This suggests that first-year students are convinced by empirical arguments, but due to the design of Segal's (2000) study, it is also plausible that the participants in her study were treating the example they were given as a generic proof. Healy and Hoyles (2000) noted that among the high school students in their study who judged an empirical argument to be valid, "many suggested in their interviews that they had made this response as a result of having 'looked through' the particular cases to the generality because the result was so obvious" (p. 411). As Segal (2000) did not use interview data in her study, the plausibility of this alternative hypothesis is unclear.

There is little research on what types of arguments advanced mathematics majors find convincing. Sowder and Harel (2003) conducted a longitudinal study in which they interviewed mathematics majors periodically to determine how their proof schemes changed over time. They present a case study of one student who still held an empirical proof scheme by the time she graduated; they also claim her case was not atypical, suggesting that many other mathematics majors would also be convinced by empirical arguments.

2. 3. What types of arguments do students consider to be proofs?

In Segal's (2000) study, students were given three empirical arguments and one deductive one and asked to judge whether they found the argument personally convincing and whether they thought the argument was a valid proof. She found that for the empirical arguments, roughly half of the participants in her sample found the arguments to be convincing but not proofs. This result, and an analogous result with first-year American students interviewed by Raman (2002), illustrate that to students, conviction and validity are not synonymous.

Harel and Sowder (1998) argue that students often hold *external proof schemes* and judge whether an argument constitutes a proof on criteria that are orthogonal to their mathematical content. Such criteria include whether the argument is written in a particular format, such as the traditional two-column format used in many high school geometry courses, whether the argument uses mathematical symbols, and even whether the argument comes from an authoritative source. Inglis and Mejia-Ramos (2009) suggest students' (and mathematicians') use of external proof schemes may be more nuanced than this. They posit a model suggesting that students may rely on external factors for

## Conviction, validity, and proof

their evaluation only when they are unable to judge the argument by its content. If correct, this suggests that students' propensity to evaluate an argument on external factors may be due, at least in part, to their inability to judge an argument on its merits.

Healy and Hoyles (2000) found that many high school students believed proofs in algebra had to contain algebraic symbols and did not think valid proofs using a narrative format would receive high marks from their teachers. In a small-scale study, Raman (2002) found similar results for first-year university students from the United States. Based on their interviews with university students in a variety of settings, Harel and Sowder (1998) concluded that many upper-level undergraduates also believe proofs must contain mathematical symbols (see Sowder and Harel (2003) for a useful case study of this). Hemmi (2006) found that some students were deeply confused by what a proof was and what its purpose was in the mathematics classroom, with much of their beliefs about the nature of proof being inferred from generalizing properties from the proofs that they had observed in their mathematics courses.

The consideration of non-mathematical factors in judging when a purported mathematical proof is valid is not limited to students. Several studies suggest that pre-service and in-service teachers exhibit this behavior as well. Martin and Harel (1989) presented valid and invalid arguments to 101 pre-service elementary teachers and found that over half of their participants accepted an invalid deductive argument as a valid proof, even though the invalid arguments used in the study contained mathematical assertions that were clearly false. Martin and Harel concluded that these participants were making judgments by the form of the argument rather than its content. Similarly, in his interviews with 16 in-service high school mathematics teachers, Knuth (2002) found that

## Conviction, validity, and proof

many accepted purported proofs by mathematical induction as valid, even though the proofs contained logical errors.

### 2. 4. Are students convinced by mathematical proofs?

Many mathematics educators argue that proof serves many roles in mathematics beyond providing conviction that a theorem is true (e.g., Bell, 1976; de Villiers, 1990; Hanna, 1991; Knuth, 2002). In an influential article, de Villiers (1990) listed several of the functions that proof serves in mathematics, including illumination, systematization, communication, and discovery. Many researchers lament that students and teachers do not see proof as satisfying these roles and view proof as a tool only used for conviction (e.g., Bell, 1976; Coe & Ruthven, 1994; Healy & Hoyles, 2000; Knuth, 2002). These broader issues of proof will not be discussed in this paper. Instead the paper will examine the extent to which, and the reasons why, mathematics majors would believe that a theorem is proven yet still have doubts that it was true.

Fischbein and Kedem (1982) found that many high school students claimed they were sure that the proof of a number theoretic statement was valid yet still desired to verify that the statement was true with particular examples. Many researchers cite this study as an example that students do not appreciate the generality or necessity of proofs (e.g., Harel & Sowder, 1998). Chazan (1993) illustrated how some high school students believed that a proof of a geometry theorem applied only to the specific diagram that was drawn but not all figures that satisfied the hypothesis of the theorem. He also reported that some students viewed proof merely as evidence in support of a statement, rather than a guarantee that the statement must be true. The extent to which university students hold such beliefs is largely unexplored, although Segal (2000) found a small number of first-

## Conviction, validity, and proof

year students judged a deductive argument to be valid but were not convinced by it and Harel and Sowder (1998) remark, in passing, that even after proving a theorem, undergraduates may still harbor some doubt that the theorem is true until they check it with examples or an authoritative source tells them that their proof is correct.

### 3. Theoretical perspective

#### 3. 1. Judgments of mathematical arguments

The primary aim of this paper is to investigate the way that undergraduates completing a transition-to-proof course make two separate judgments on ten mathematical arguments. The first judgment asks students how convincing they find the mathematical argument. It is assumed that this judgment involves evaluating the extent to which the argument is *personally* convincing and is related to the ascertaining aspect of Harel and Sowder's (1998, 2007) proof schemes. The second judgment asks the students if they believe the argument constitutes a valid proof. Although it is not possible to determine how the participants are making this judgment, the assumption is that participants are evaluating whether the arguments would be sanctioned as proofs by the broader mathematical community. As the participants most likely do not yet feel that they are full members of this community (e.g., Hemmi, 2006), a further assumption is that this judgment is based on whether the participants believe *other mathematicians*, including the professor of their transition-to-proof course, would find the arguments to be acceptable. Dreyfus (1999) notes that whether an argument constitutes a proof is a sociomathematical norm. Hemmi (2006) illustrates how some students infer what constitutes a proof from the exemplars that professors provide for them in their lectures.

## Conviction, validity, and proof

Consequently, participants' judgments on what types of arguments constitute proofs are likely to be influenced by the sociomathematical norms and the proofs that are presented in their proof-oriented courses.

A central hypothesis in this paper is that participants may consider different factors when determining whether an argument is convincing or whether it constitutes a proof. The former judgment is assumed to be dependent upon personal criteria. The latter might be influenced by their experience as students in proof-oriented courses and may depend upon such factors as comments from their professors about what factors a proof needed to have and similarities between the arguments they were reading and the proofs they had seen in their courses. Participants may defer to their professors' judgment on what constitutes a proof as the professors are expert members of the mathematical community.

One issue addressed in this paper are the reasons for inconsistencies in participants' evaluations of the arguments—i.e., arguments they found to be completely convincing but not proofs and arguments they judged to be proofs but not completely convincing. There is disagreement in the mathematics education community as to whether students should view all convincing arguments to be proofs (and vice versa) and whether mathematicians do the same. For instance, Mason, Burton, and Stacy (1981) define the act of proving as convincing oneself, a friend, and then an enemy, and Balacheff (1987) defines a proof as an argument that convinces a given community at a given time. Harel and Sowder (1998) contend that contemporary mathematicians hold axiomatic analytical proof schemes, implying that deductive proofs in axiomatized theories, but not other types of evidence, remove all doubt about the truth of theorems.

## Conviction, validity, and proof

However, others question the claim that mathematicians obtain conviction from proof alone. Tall (1989) noted that not all convincing arguments would be classified as proof. de Villiers (1990) avers that mathematicians are usually convinced a claim is true before trying to prove it and, in some cases, mathematicians may read a proof but not obtain complete conviction.

### 3. 2. Empirical evidence

A universal statement is one that asserts that all elements of a given set have a specific property. For instance, the claim that “the square of every odd number is odd” is a universal statement averring that every element of the set of odd numbers has the property that its square is odd. There are (at least) three ways that empirical evidence might convince an individual that the assertion is true. First, an individual might check if the assertion holds in one or several instances, obtaining no information from this process other than a verification that the assertion holds true in these instances. This is sometimes sufficient to convince an individual that the assertion will be true in all relevant cases. Balacheff (1988) refers to this type of verification as *naïve empiricism*.

At a more sophisticated level, the process of checking whether the assertion is true with specific examples may provide more information than simply providing verification that the assertion is true in those cases. It may suggest a reason why the assertion is true. To illustrate this distinction, imagine a student who was asked whether every natural number  $n$  had the property that  $\frac{n^3 - n + 12}{6}$  was composite. The first six elements of this sequence are 6, 9, 18, 36, 66, 111. If the student simply verified that these numbers were composite (perhaps using a calculator that could check the primality of each given integer) and then concluded that this would be true for all natural numbers,

## Conviction, validity, and proof

this would be an instance of naïve empiricism. However, in the process of checking for primality, a student might realize that each integer in this sequence was divisible by 3 (which would ensure it was composite if the integer was greater than 3). This could provide greater conviction that the assertion was true because it suggested a reason for why it was true. Pedemonte (2007) refers to this type of verification as *abductive reasoning*; she argued that this is preferable to naïve empiricism because it reduces the distance between verification and proof. More sophisticated still is the idea that verification of an example can serve as a *generic proof* (Mason & Pimm, 1985; Rowland, 2001). When this occurs, an individual recognizes that the *process* of verifying the assertion for a particular assertion can be generalized and applied to any assertion.

In the mathematics community, arguments displaying naïve empiricism, abductive reasoning, or generic proofs all would not receive the status of being a mathematical proof (although all are undoubtedly useful in mathematical reasoning and the construction of proofs). The normative view in the mathematics education community is that students should not obtain full conviction in the truth of an assertion by naïve empiricism or abductive reasoning, since they cannot be certain that the trends that they noticed would continue to hold for all elements under consideration. However, many argue that generic proofs can provide legitimate conviction<sup>2</sup>. It is an interesting and open question as to whether, and under what circumstances, mathematicians might gain complete conviction from each of these types of reasoning.

When a student claim to be convinced by an empirical argument, it can be difficult to determine if this student is using naïve empiricism, abductive reasoning, or a

---

<sup>2</sup> Harel (2001) notes that generic examples are instances of the desirable transformational proof scheme, not the empirical proof scheme. Rowland (2001) suggests that it may be appropriate to use generic proofs in lieu of formal proofs because they can be equally convincing, or perhaps more so.

## Conviction, validity, and proof

generic proof to obtain this conviction. As noted earlier, when Healy and Hoyles (2000) presented students with what would seem to be an argument based on naïve empiricism, some who were convinced by the argument claimed that they “looked through” the particular examples to generality; in other words, they were able to view the particular examples as generic proofs. Alternatively, Mason and Pimm (1985) and Rowland (2001) warn that students might fail to see the generality in a generic proof that they read and believe the generic proof only verifies the claim for a single example. This introduces a methodological difficulty in measuring what types of arguments convince students that will be discussed in the next section.

### 3. 3. Diagrammatic evidence

Individuals may increase their conviction in the veracity of an assertion by the inspection of diagrams. Similar to examples, there are many ways in which a diagram may provide conviction. I describe four ways that an individual may interact with a diagram to obtain this conviction<sup>3</sup>:

- (a) The individual recognizes how the diagram relates to the statement that it purports to verify.
- (b) The individual believes that the measurements, or other relevant features of the diagram, are accurate.
- (c) For universal statements, the individual recognizes the generality of the diagram. That is, the individual understands how the diagram can be deformed or extended to account for all instances to which the universal statement applies.

---

<sup>3</sup> Of course individuals may have their own idiosyncratic reasons for gaining conviction from a diagram. I am simply describing here conditions that I believe to be common and important.

## Conviction, validity, and proof

(d) The individual sees a way that the information in the diagram can be translated into a formal mathematical proof.

Each of these points is illustrated by considering the following statement:  $4x^3 - x^4 = 30$  has no solutions (adapted from Selden, Selden, & Mason, 1989). Suppose that upon reading this assertion, a student sketches the graphs of  $f(x) = 4x^3 - x^4$  and  $g(x) = 30$  and observes that  $f(x)$  and  $g(x)$  do not intersect.

Condition (a) would be satisfied if the participant understood that the number of solutions to an equation of the form  $f(x) = g(x)$  was equal to the number of times these two graphs intersected. Condition (b) would be satisfied if the participant was convinced both that the graph was accurate and the two graphs really did not intersect. This might be an issue if, for instance, the student used a graphing calculator with poor resolution. In some situations, such as comparing the areas of geometrical figures, making such judgments can be difficult. The individual might be convinced by the accuracy of a diagram either for *perceptual reasons* (e.g., an area looks bigger) or for *mathematical reasons* (e.g., the area must be bigger due to its mathematical properties).

In the example above, condition (c) would be satisfied if the participant believed that  $f(x)$  and  $g(x)$  not only failed to intersect in the range that the graph was inspected, but for all values of  $x$ . The individual may do this for *mathematical reasons* (e.g., citing the graphical behavior of quartic functions) or for *inductive reasons* (i.e., the trends observed for some values of  $x$  will be true for all values of  $x$ ). In general, one way that (c) might be satisfied is by what Simon (1996) refers to as *transformational reasoning*. Condition (d) would be satisfied if the individual saw how he or she could translate what was present in the graph into the language of formal proof. For instance, in our example, the participant

## Conviction, validity, and proof

might realize that it would be relatively easy to use calculus to find the global maximum of  $f(x)$  and show that  $f(x) < 30$  for all  $x$ .

There is currently a debate in the philosophical community as to whether diagrammatic arguments should, or do, constitute an acceptable form of proof. Inglis and Mejia-Ramos (2008b) assert that the common view of the relationship between diagrams and proof is that “pictures may be useful heuristic tools which suggest ways of understanding proofs but that they are nevertheless inappropriate when it comes to providing unequivocal reliable evidence to support a mathematical claim, let alone providing a proof”<sup>4</sup> (p. 124). Other philosophers (e.g., Brown, 1999; Giaquinto, 2005; Kulpa, 2009) and mathematics educators (e.g., Eisenberg & Dreyfus, 1991) reject such a viewpoint and argue that some diagrammatic arguments (ones that satisfy conditions (a), (b), and (c)) should constitute an acceptable form of mathematical proof<sup>5</sup>. Supporting this position is Herbst’s (2004) observation that both historically and in classroom practice, mathematicians have implicitly relied on perceptual reasoning to draw mathematical inferences (e.g., judging whether a point is in between two points on a line or in the interior of a geometric figure). Rav (1999) also argues that in some new fields, such as knot theory, using diagrams in proofs to draw inferences is common practice.

## 4. The research study

### 4. 1. Participants

Twenty-eight mathematics majors who had recently completed a transition-to-proof course agreed to participate in this study. All were mathematics majors in their

---

<sup>4</sup> Inglis and Mejia-Ramos (2008b) do not endorse this common view in their paper.

<sup>5</sup> In my personal judgment, conditions (a), (b), and (c) would be useful criteria for determining if a diagrammatic argument should be classified as an *informal proof*.

## Conviction, validity, and proof

sophomore or junior year. These students volunteered to participate in the study after an e-mail solicitation was sent to all students who had completed the transition-to-proof course in their previous semester. The researchers had no involvement with any aspect of this course. Each student was paid a small fee for his or her participation. The grades that these participants received in their transition-to-proof course ranged from A to D, with the median grade being a B. The median grade for the course on the whole ranged from B to C (depending on the instructor), so the participants in this study performed above the department median.

### 4. 2. Materials

To generate the materials used for this study, my research team and I generated 20 mathematical arguments. These arguments were tested in pilot studies with mathematics majors. Arguments that were difficult for the participants to understand because they lacked familiarity with the concepts involved were removed from the argument pool. This process was used to narrow down the pool of arguments to ten. The ten arguments used for this study are presented in the Appendix.

The arguments varied along several dimensions, including the type of argument (empirical, diagrammatic, and deductive), the format of the argument (narrative and symbolic), and the mathematical content (high school algebra, number theory, and calculus). Three of the seven deductive arguments were valid proofs (arguments 1, 3, and 4) and four contained logical errors (arguments 7, 8, 9, and 10). A summary of the attributes that each argument has is presented in Table 1.

**\*\*\* Insert Table 1 Here \*\*\***

The most significant dimension in this study was the *type of argument*. Here, the type of argument refers to the nature of the evidence contained within the argument that is supposed to convince the reader that the claim is true. Three types of arguments were considered in this study—empirical, diagrammatic, and deductive—but this is not an exhaustive list. (For instance, an argument based on an appeal to authority could also have been included). These three types of arguments were chosen because they have been of particular interest to mathematics educators in recent years. Note that argument 7, 8, 9, and 10 were *invalid deductive arguments* in the sense that the conclusion of the arguments *purported* to be the logical consequence of reasonable assumptions, but in fact were not. Note that one could analogously produce *flawed empirical arguments* (e.g., arguments based on examples with calculation errors) or *flawed diagrammatic arguments* (e.g., arguments with inappropriate or misleading diagrams), but none were included in this study.

#### 4. 3. Procedure

Each participant met individually with the author for a task-based interview. Participants were told they would be presented with ten arguments, one at a time, and that they would be asked to make three judgments on each argument. First, they were asked to rate on a five-point scale the extent to which they felt they understood the argument, where a score of 5 indicated that “I understand this argument completely”. Second, they were asked to rate how convinced they were by the argument using a five-point scale, where a score of 5 indicated that “I am completely convinced by this argument”. Third, they were asked to decide whether the argument was a proof. They were given four choices: 1) the argument was a rigorous proof; 2) it was a non-rigorous proof; 3) it did

## Conviction, validity, and proof

not meet the standards of a proof; or 4) they were not sure because they did not fully understand the argument. They were also permitted to opt for “other” if they did not feel comfortable with these four choices.

Arguments 1 and 2 were the first two arguments presented to all the participants. The remaining arguments were presented to the participants in a randomized order. The decision to provide arguments 1 and 2 first for all participants was based on the results from pilot studies in which participants would express confusion when they saw the same assertion justified in two different ways. When presented with argument 2, participants were informed that both argument 1 and argument 2 supported the same assertion and that this would be the only instance in the study in which this occurred.

It was emphasized to the participants that their judgments on the arguments should come from what was contained in the argument, and not from their knowledge of whether the claim being proven was true or false. Specifically, participants were told that even in cases where they knew the claim was true, if they found the argument to be unconvincing, then they should rate it as such. Participants were also informed that some of the arguments would be “good arguments” while others would be “flawed”. The participants were told that they should spend as much time as they liked while reading the arguments.

Participants were given each argument individually and asked to “think aloud” as they read the argument and made their judgments using the verbal protocol procedures described in Ericsson and Simon (1993). An analysis of what the participants attended to while reading the arguments is beyond the scope of this paper, but preliminary analysis on this topic can be found in Weber (2009). If participants claimed they did not

## Conviction, validity, and proof

understand an argument or did not find it fully convincing (i.e., gave a mark less than a five for these two judgments), or did not find the argument to be a proof, they were asked why they gave it that mark.

After reading all ten arguments, participants were asked a series of open-ended questions about how they read arguments (e.g., “What are some of the things that you do when you read a mathematical argument?”) or about their perceptions of mathematical arguments (e.g., “What do you think makes a good mathematical argument?”). They were also asked to rate, with justification, their favorite argument, their second favorite argument, and their least favorite argument.

### 4. 4. Analysis

Each participant’s justification for why an argument was not fully convincing or not a proof was assigned to a group of similar types of justification using the constant-comparative method (Strauss & Corbin, 1990). Each justification was given an initial description. Similar justifications were grouped together and given preliminary category names and definitions. New episodes were placed into existing categories when appropriate, but also used to create new categories or modify the names or definitions of existing categories. This process continued until a set of categories was formed that were grounded to fit the available data.

### 4. 5. The empirical argument

Argument 5, presented in the Appendix, is an argument that purports to demonstrate that every even number greater than two can be written as the sum of two primes; in the argument, it is verified that the claim holds for the even numbers between 4 and 26. No calculations are shown to students and there are no significant patterns in

the calculations that suggest a reason for why this assertion is true. This argument was used to assess whether the participants would find an argument based on naïve empiricism to be convincing. In analyzing participants' evaluations of this argument, for favorable evaluations, I coded for any instances in which there was evidence that participants noticed a pattern in the data or used the data to conjecture a reason for why the assertion is true or viewing the examples to be generic. Also, I coded any instances in which participants noticed that they were being shown a proof of an open mathematical conjecture.

#### 4. 6. Diagrammatic arguments

Arguments 2 and 6, presented in the Appendix, are diagrammatic arguments in the sense that critical deductions in the proof were supported solely by appeals to perceptual features of the diagram. Argument 2 justifies the identity  $(a + b)^2 = a^2 + 2ab + b^2$  using an area model of multiplication. It does not consider the scope of  $a$  and  $b$  or the fact that such a model would be difficult to interpret if  $a$  or  $b$  or both variables were negative (although the scope of  $a$  and  $b$  were not defined in the problem statement).

Argument 6 justifies that  $\int_0^{\infty} \frac{1}{x} \sin x dx > 0$  by an appeal to the graph of  $f(x) = \frac{1}{x} \sin x$ , noting that the graph is a sequence of positive and negative areas, and the positive area is always bigger than the negative area immediately to its left; three positive and three negative regions were displayed in the range of the graph. This appeal is entirely perceptual; no deductive justification is given for why the positive areas should be bigger, nor is there a justification for why this trend should continue. Each proof contains a paragraph explaining how the diagram relates to the proof (condition (a) in section 3.3). When analyzing the data, it was noted when students made comments to items (b), (c),

and (d) in section 3.3—that is, comments related to the accuracy of the diagrams, the extent to which the diagrams can be transformed or extended to cover all possible cases, and if they could see how to use the diagrammatic evidence to construct a formal proof.

## 5. Results

A frequency summary of the main results from this study is presented in Table 2. These results, tabulated by type of argument, are summarized in Table 3. Table 4 presents the number of times that participants judged an argument to be completely convincing (i.e., they gave the argument a score of 5 in terms of being convincing, where a 5 represented being “completely convinced”) but not a proof or a proof but not completely convincing (i.e., a score of 4 or less). These instances will be highlighted in the discussion to illustrate how the participants sometimes used different criteria in determining whether an argument is convincing and whether an argument is a proof.

**\*\*\* Insert Tables 2, 3, 4 About Here \*\*\***

### 5. 1. Evaluations of the empirical argument

#### 5. 1. 1. Empirical arguments and conviction

In general, the participants did not find the empirical argument to be convincing. Only one of the 28 participants found argument 5 completely convincing; 22 of the participants gave the argument a score of 3 or less in terms of being convincing. The average score for conviction for this argument was 2.61, easily the lowest average score of the ten arguments in this study. Only two of the 28 participants judged the argument to be a proof. Finally, at the end of the interview, participants were asked to name their favorite argument, their second favorite argument, and their least favorite argument. None of the

## Conviction, validity, and proof

participants listed argument 5 as one of their two favorite arguments while a plurality of the participants, 11, rated argument 5 as their least favorite argument.

When asked why they did not find the argument fully convincing, 26 participants responded by articulating the mathematical limitations of empirical reasoning. Two representative responses are provided below:

P11: I am not convinced by this argument. The argument is wrong. It's a terrible argument, although it kind of shows a pattern that does make sense, so say I'm partially convinced by the argument. But I would say this is not at all a rigorous proof, or proof at all. Does not meet the standards of a proof.

I: Why would you say it's not a proof?

P11: Because this only shows that 26, doesn't show any above that. You can't just make an assumption based on the first 26 numbers. Not even 26 numbers.

P2: I understood what they were saying. I'm not convinced since there's a lot more prime numbers than the ones that they presented. So it's not really a proof because it's a set that goes a lot larger than what they gave.

In these excerpts, P2 and P11 acknowledge that checking that a claim about an infinite set holds in a small number of cases (14 for this argument) is not sufficient because it might not apply to all elements of the set.

One possible explanation for this data is that the participants recognized Goldbach's conjecture as a well-known open mathematical problem and consequently believed that the argument in support of it was likely incorrect. However, no participant made a comment of this type, either when reading the argument or providing a justification. On the contrary, six of the participants actually remarked that they believed the assertion was true because they recalled the professor proving this theorem in class. No participant indicated that they saw a pattern in this argument or viewed the argument

generically. In summary, these data suggest that the participants were not using naïve empiricism (as defined in section 3.2) to obtain conviction.

#### 5. 1. 2. Empirical arguments and validity

As Table 2 and Table 3 illustrate, only two participants judged argument 5 to be a non-rigorous proof while the remaining 26 stated the argument did not meet the standards of a proof. Most participants responded to why they found the argument unconvincing and not meeting the standards of proof at the same time. These comments were reviewed in the previous sub-section. The participants' judgments on both conviction and validity, as well as their comments on why they made these judgments, do not support the claim that undergraduates find empirical arguments to be an acceptable form of proof.

#### 5. 1. 3. A completely convincing empirical argument that was not a proof

As almost no participants found argument 5 to be either fully convincing or an acceptable proof, there were not many instances where a participant found the argument to be convincing but not a proof, or vice versa. However, as Table 4 illustrates, the one student who did find the argument completely convincing (i.e., gave it a score of 5 for being convincing) did not judge the argument to be a proof. P6 judged argument 5 to be completely convincing, arguing that “the table kind of showed me why it works”, since you could find a pair of primes that add to  $(n+2)$  by taking two primes that add to  $n$  and adding two to one of these primes<sup>6</sup>. However, P6 believed the argument was not a proof because it did not have equations, remarking:

---

<sup>6</sup> This does not work, of course. As a counterexample, 13 and 19 are two primes that add to 32. But adding two to either will produce a composite number. Hence one cannot follow P6's procedure to use 13+19 to find two primes adding to 34.

## Conviction, validity, and proof

P6: I think that like they should of like, like kind of did like step by step into showing them how like it can be done. Like maybe with like  $2n-1$  times by  $n$  or something, I mean plus  $n$ . Like, like an equation. [...] I just think they should just showed like an equation therefore people can plug in the numbers.

Hence, although P6 was able to gain conviction from the argument, he believed it needed to contain an equation to be a proof (a belief he expressed elsewhere in the interview as well).

### 5. 2. The diagrammatic arguments.

#### 5. 2. 1. Diagrammatic arguments and conviction

Argument 2 presents a visual area-based model that illustrates why  $(a+b)^2 = a^2 + 2ab + b^2$ . Twenty-two of the 28 participants rated the argument as fully convincing. None of these 22 students referenced the accuracy or generality of this diagram, either when reading the argument or making their evaluation, and none discussed how this argument could be written as a formal proof.

Of the six who did not find the argument fully convincing, two noted that the diagram might not be sufficiently general to cover every possible  $a$  and  $b$ . One participant remarked that the diagram would only apply to the positive numbers while another noted that the diagram would not be applicable to the complex numbers. Three of the other participants who did not find the argument fully convincing also cited the diagram as a reason for their judgment but did not give a specific explanation for why diagrams were not convincing, making comments such as “Usually diagrams, from what I’ve heard, diagrams don’t constitute a proof”. Half the participants, 14 of 28, listed argument 2 as their favorite argument, with most citing its explanatory power or its novelty as the reason for their positive opinion of it.

Argument 6 uses a graph in a critical way to justify that  $\int_0^{\infty} \frac{1}{x} \sin x dx > 0$ . Eleven of the 28 participants found the argument fully convincing. None of the 11 participants who found the argument fully convincing provided a justification for why the negative areas were smaller than the corresponding positive areas; they appeared convinced of this on the appearance of the graph alone. Two participants mentioned that the graph only applied to a limited range of the function, but provided mathematical reasons for why this should not be an issue. P15 cited the behavior of the function, noting the periodicity of the sine function; P5 argued that the end behavior of the graph would not matter so much, “from the graph given... when the function is getting really big, the graph is going to be really flat at the end. Therefore it should always be positive because from zero to  $\pi$ , you have the biggest jump. And the biggest jump is going to cover everything else”. One participant, again P5, mentioned that they could see how this argument could be mapped

## Conviction, validity, and proof

to a more formal proof, arguing (incorrectly) that  $f(x) = \frac{1}{x} \sin x$  could easily be integrated by parts.

Three of the 17 participants who rated the argument as not fully convincing did so because they did not fully understand it. Four of the participants questioned the accuracy of the graph; two mentioned that they could not be convinced from a pictorial argument because the picture might be misleading and two claimed they had difficulty determining if the third positive region was actually bigger than the third negative region. Five participants were not fully convinced of the argument since the graph in the argument was of limited range. The remaining five participants rejected the argument because it had a graph, but did not give a specific reason for why a graphical argument could not be convincing.

### 5. 2. 2. Diagrammatic arguments and validity

Twenty-four of the 28 participants judged argument 2 to be a proof, although 14 of these 24 participants did not think it was rigorous. Of the 14 students who did not find the proof to be rigorous, eight cited the use of a diagram as a reason the argument was fully rigorous.

Twelve of the 28 participants found argument 6, which was based on a graph, to be a valid mathematical proof, although only three students judged it to be rigorous. 14 participants said that argument 6 did not meet the standards of a mathematical proof, with nine of these participants citing the use of a graph as a reason. In summary, most of the participants believe at least some arguments that rely on diagrams can constitute mathematical proofs, although many classified these arguments as non-rigorous proofs,

although nine participants judged argument 6 not to be a proof because of the presence of a graph.

5. 2. 3. Completely convincing diagrammatic arguments that were not proofs

As Table 4 illustrates, there were five instances in which a participant found argument 6 to be fully convincing but did not consider it to be a valid proof. In all cases, these participants cited the presence of the graph as the reason for this judgment. Four of the participants stated that they had been told that proofs were not allowed to use graphs. Consider P5's evaluations of argument 6 below.

P5: I guess I understood the argument. Am I convinced? Well, I'm convinced by the graph. So I'll say five. Would you consider it a mathematical proof? I wouldn't consider it a proof. I would say no, I do not think this argument meets the standards of a proof.

I: Why not?

P5: Because most of proofs I would, *they don't really allow us to use graph*. And this, *I'm basing this whole proof off of a graph which is not, I would say, professional*. Usually proof, we have to find a value such that for all these values, this would happen. *So, that's what I've been learning in class. We're not allowed to draw any pictures*. And I'm basing this whole thing off of a picture so I would say this is not proof. I am pretty sure they have a better way of proving it using words. (Italics are my emphasis)

In the first italicized excerpt, P5 cites unspecified authorities, presumably his instructors, who do not permit the use of a graph in a proof. This is reiterated in the third italicized excerpt. As the second italicized excerpt illustrates, he finds the use of a graph to be “not professional”, indicating that he believes that it does not conform to the norms of the mathematical community, even though he indicated earlier that he found the argument to be completely convincing due to the presence of the graph.

One student, P10, also believed that graphs were not permitted in mathematical arguments, although he was less certain of this. As the excerpt below illustrates, P10 justifies his response by saying that he usually doesn't see graphs in mathematical proofs.

P10: Ok, it's definitely easy to follow. Do you feel you understood the argument? Yes, completely. Are you convinced? I am convinced. Would you consider this argument to be a mathematical proof? Hmm... that is a toughie. I mean it certainly doesn't smell like a mathematical proof in that it is very intuitive. You don't see many graphs in mathematical proofs. I'm going to say no, I don't think it meets the standards of a proof. Of course, you want to know why it doesn't meet the standards of a proof. I'm trying to preempt your question. That's hard to say. Could be I just haven't seen many proofs involving areas under integrals.

Like the other students, P10 is basing his judgment upon his experience in the mathematical community, rather than on what he finds to be personally convincing.

### 5. 3. The deductive arguments.

#### 5. 3. 1. Deductive arguments and conviction

Deductive arguments made up the majority of arguments in the study. Seven of the arguments were deductive and three were valid proofs. Every participant judged at least two of the deductive arguments to be fully convincing and no participant commented that they did not find deductive reasoning to be convincing. Further, 25 of the 28 participants rated at least one deductive argument with a score of 3 or less in terms of being convincing, suggesting that these participants did not find *all* deductive arguments to be convincing and they were attending to the mathematical content of the deductive arguments that they were reading, at least some of the time. This point is discussed in more detail in the next sub-section. In summary, the participants found valid deductive

## Conviction, validity, and proof

arguments to be convincing but, at least in some cases, attended to their mathematical content in forming these evaluations.

### 5. 3. 2. The form of deductive arguments and validity

Twenty-three of the 28 participants judged at least one of the seven deductive arguments in this study to not meet the standards of a mathematical proof. Of these 23 participants, 20 cited a reason pertaining to the mathematical content of the proof—that is, they rejected the proof because it contained an assertion that was false, not appropriate for a proof (e.g., beginning a proof by assuming the conclusion), or did not follow validly from previous assertions. This suggests that for the majority of the participants, a deductive argument that incorporates symbols is not, by itself, sufficient cause to judge the argument to be a proof. The mathematical content of the argument also influenced how they evaluated it, at least some of the time. Of course, it would be inappropriate to conclude that participants who judged all seven arguments to be proofs did so solely because of the form of these arguments; it is possible that they simply overlooked the flaws in the invalid proofs.

At several points in this study, participants commented that the proofs they were observing were unlike the types of proofs that they had seen before. Four participants commented that argument 1 did not seem like a proof since it was simply performing algebraic manipulations on an equation. For instance, after reading the argument, P9 remarked, “I don’t really see a proof. It’s more like just doing some kind of math... it’s just doing math here. I don’t feel like it’s a proof for some reason”. Similarly, P12 commented:

## Conviction, validity, and proof

P12: I mean, I forget the way these proofs are supposed to be presented, but like you can see that they're just expanding the math. There should be like more, I don't know, like theory behind it, I guess.

I: You mean, like an explanation for the steps?

P12: Yeah, something like that.

I: If they added, "I'm using the distributive law here" or something like that, would that make it better?

P12: Yeah, yeah.

Nonetheless, all four participants judged the argument to be a convincing, non-rigorous proof. In these cases, it is somewhat debatable if students are arguing about the form of the argument (e.g., a proof cannot just be "doing math" on equations) or the content (e.g., the underlying mathematical principles used in a proof should be specified). Four participants commented that they believed argument 3 did not seem right since the claim was about the integers and therefore should have been proved using mathematical induction. One of these four participants claimed that argument 3 was not a proof, two thought it was a proof but did not find it fully convincing, and one rated it as a proof that was fully convincing. In summary, comments about the format of the proof were relatively rare (only 8 such comments in the 280 proofs the participants collectively read) and, in the majority of cases, did not affect the participant's judgment.

### 5. 3. 3. Participants' acceptance of invalid deductive arguments

Arguments 7, 8, 9, and 10 were deductive arguments that contained logical flaws (see Table 1). As Table 2 illustrates, participants collectively judged these arguments to be proofs (this includes both rigorous and non-rigorous proofs) in 64 out of 106 instances, or 60% of the time. (Instances in which participants selected "not sure" or "other" were not included in this tally). A complete discussion of why the participants

## Conviction, validity, and proof

made these errors is beyond the scope of this paper; a preliminary analysis of this topic can be found in Weber (2009). In general, participants who made these errors did so because: (a) they did not check if the assumptions used in a proof were appropriate (a finding also reported in Selden and Selden (2003)), (b) they focused on the correctness of the assertions within the proof but not on what mathematical principles were used to deduce new assertions from previous ones (a finding also reported in Alcock and Weber (2005)), (c) limitations in their own mathematical knowledge prevented them from spotting flaws in the arguments they were reading (e.g., six participants believed that increasing functions necessarily diverge to infinity as  $x$  approaches infinity; if true, argument 9 would be valid), and (d) simple oversight.

Although participants rarely rejected an argument based on its appearance, four participants (three for argument 9, one for argument 10) commented that the form of the argument that they were reading gave them a false sense of confidence that the entire argument would be correct and led them to overlook errors in the argument. For instance, P20 states that he incorrectly believed the last line of argument 9 followed validly from previous assertions because the proof up until that point had been correct.

P20: The therefore. Yeah, like, I was reading all of [the lines of the proof] and I got it. But then, like, all of a sudden like this popped up. And I was reading it and like I believe like all of it was true and then the therefore popped up. So I believe that it was a part of it so then I thought it was true.

It is possible that other students judged arguments 9 and 10 to be proofs for similar reasons. Nonetheless, comments of this type were not frequent. For the most part, the form of the argument did not appear to influence whether the participants believed that the argument was a proof.

5. 3. 4. Completely convincing valid arguments that were not proofs

As Table 4 illustrates, three participants considered argument 3 to be fully convincing but not a proof. Two participants argued that more detail needed to be provided. This is illustrated with the excerpt below:

P10: [reading argument 3] Since it is even and divisible by 3, then it's divisible by 6. [commenting] Yes. The fact that it's even means it is divisible by 2... it's divisible by 6... [inaudible] multiple of 6. I would like to see that made explicit. Do you feel that you understood the argument that was presented? I completely understand that. Are you convinced by the argument? I am completely convinced because I am able to fill some of the gaps. I don't consider this to be a fully rigorous proof because a few things should have been made explicit. I am going to say it does not meet the standards of a proof because of the things I mentioned. It doesn't explicitly say... although if you're dealing with a more advanced mathematician, you obviously leave out more stuff and expect them to.

Like the mathematicians in Weber's (2008) proof validation study, P10 appears to argue that the level of detail that needs to be provided for an argument to be sanctioned as a proof depends upon the audience of the proof and community in which it is couched. He felt that with argument 3, more argument needed to be provided for the last step of the proof.

One participant judged the last argument to be valid but not a proof, but was unable to articulate a reason for her judgment.

5. 4. Participants' reasons for judging an argument to be a proof but not being fully convinced by it

As Table 4 illustrates, in the 150 instances in which a participant found an argument to be fully convincing (i.e., a score of 5 for conviction), they judged the

## Conviction, validity, and proof

argument to be a proof in 139 of those instances, or 92.7% of the time. In 10 instances, or 6.7% of the time, they evaluated the argument to not be a proof. In one instance, the participant was unsure.

There were a total of 130 instances in which participants indicated that they were not completely convinced by an argument (i.e., a score of 4 or less for conviction). In 80 of these instances (61.5% of the time), the participants also believed the argument did not meet the standards of a proof. In 11 instances, the participants either were unsure if the argument was a proof (usually because they did not fully understand it) or coded the argument as “other” when asked if it was a proof. The analysis in this section concentrates on the 39 instances when participants found the argument to be a proof, but not fully convincing. This represents 30.0% of the instances in which participants were not completely convinced of the argument. Twenty-one of the 28 participants made such a judgment.

In 10 of these instances, the participant was not sure about a particular step within the proof. For instance, when reading argument 3 in support of the statement,  $n^3 - n$  was divisible by 6, P1 responded:

P1: [Reading the proof] So since  $n$  cubed minus  $n$  is even and divisible by 3,  $n$  cubed minus  $n$  is divisible by 6. [Commenting] I don't remember all of the rules for everything for all of those division things, so that's holding me back right now. So I guess I understand the argument for the most part. And I guess I can say that I am pretty much convinced [...] I forgot what it means, the last part, like what numbers are divisible by 3, divisible by 6, I forgot those rules.

P1 rated this argument as a 4 in terms of being convincing, because he was unsure if being even and a multiple of 3 necessarily implied that the number was divisible by 6.

(The implicit warrant in this conditional statement is that a number divisible by  $m$  and  $n$

will necessarily be divisible by  $mn$  if  $m$  and  $n$  are coprime). Hence in this case, and other cases like it, the participants are not saying the arguments are definitely proofs, but are proofs on the condition that a particular assertion within the proof is correct and follows validly from previous assertions.

While such judgments are not inherently irrational, they also are not consistent with the behavior of professional mathematicians. In a study in which I asked mathematicians to judge the validity of purported proofs, if the mathematicians were unsure about the validity of a single statement within the proof, they would not hazard a guess as to whether the proof was valid, instead saying they were unsure (Weber, 2008). Furthermore, the participants in this study usually spent under two minutes studying each argument before evaluating it while the mathematicians in Weber (2008) sometimes spent as long as ten minutes studying a single assertion within the proof<sup>7</sup>. The point of noting these differences is to argue that when students are not completely convinced of a proof they judge to be conditionally valid, this may not indicate inaccurate beliefs about what a proof is or what a proof accomplishes, but about what their responsibilities are when they are reading a proof and how much time they should spend sorting out their difficulties.

In 13 instances, the participants could not find anything wrong with a specific assertion within a proof, but had difficulty following the general flow of the argument. This happened fairly often with argument 4 and is illustrated with P19's response below:

---

<sup>7</sup> The proofs explored by these mathematicians were much more sophisticated. However, the point is that mathematicians would spend a significant amount of time studying a single assertion, while the participants in this study would spend little time on this task.

## Conviction, validity, and proof

P19: For convincing, hmm, I'll give it a 4. Is it a proof? Yes, they've shown that it [referring to  $4x^3 - x^4$ ] has a maximum of 27 and is decreasing after that so it could never be equal to 30. Yes, it's a rigorous proof.

I: What didn't you find convincing about the argument?

P19: The whole, I didn't remember how you found global maximums. Each of the steps made sense though. I think it's right, but I'm not sure.

This seems similar to the situation in which participants judged an argument to be a proof on the condition that a particular assertion in the proof followed validly, except they were unable to pinpoint a specific point that gave them difficulty.

There were six instances in which a participant thought each step in the argument was correct and therefore thought the argument in its entirety was a correct proof.

Nonetheless, they did not find the argument personally convincing. Consider P9's response when reading argument 3:

P9: It's just not right, I think [...] I couldn't find something wrong here but for some reason it doesn't convince me. How about I put as a three? Neutral?

I: Okay. Even though you can't find anything wrong, you just don't find it convincing?

P9: I have to think about this one [...] I think it is right, this proof. But I don't know. Let me just put I consider this argument to be a proof although not fully rigorous.

The remaining 10 responses could not be grouped into large categories. Two participants judged argument 3 to be a proof, but found it unconvincing because they believed the proof should proceed by mathematical induction. Two participants rated argument 7 as a proof, but found it unconvincing because it began by assuming the conclusion and deducing the hypotheses. (This should invalidate the proof, but the students did not think of this as a significant flaw. One dismissed it as "formatting"). The other five responses were either idiosyncratic or unclear.

## 6. Discussion

### 6. 1. The empirical argument

Harel and Sowder (2007) argue that a significant goal of mathematics instruction should be to have students come to share the same proof schemes held by mathematicians. While there is ample evidence that elementary and high school students hold empirical, inductive proof schemes (i.e. are convinced of a general assertion because it holds for one or several examples), the extent that this is true with advanced mathematics majors has not been systematically assessed. The results from this paper suggest that the mathematics majors in this study provided evidence that they did not hold empirical proof schemes, at least in the sense of naïve empiricism. Nearly all participants in this study found the empirical argument (argument 5) to be an unconvincing argument that does not qualify as a proof, clearly articulating the limitations of empirical reasoning as a justification for these evaluations. As participants read only a single empirical argument, more research would be needed to confirm the reliability of this finding<sup>8</sup>.

Below, I list four hypotheses for why the mathematics majors in my study did not provide evidence of holding empirical proof schemes while other studies have reached the opposite conclusion with elementary and high school students. First, I inferred students' proof schemes from how they evaluated the arguments of others, while others sometimes did so by examining the arguments that students produced. These different tasks may elicit different proof schemes and, as I argued earlier in the paper, it may be inappropriate to infer proof schemes from the invalid proofs that students produce.

---

<sup>8</sup> It is possible, for instance, that the table in Argument 5 suggested to participants that the proof was invalid or that participants would have been convinced if a large number of examples were verified.

## Conviction, validity, and proof

Second, when evaluating the arguments of others in previous studies, students may have treated the examples that they were shown as generic proofs (as observed in Healy and Hoyle's (2000) study). However, the empirical argument in this paper does not lend itself to this type of generalization. These first two hypotheses essentially posit that the empirical proofs schemes held by high school and university students may be less robust than researchers commonly believe.

The third possibility is that the transition-to-proof course that students completed may have led them to revise their proof schemes. The data from this study do not allow a test of this hypothesis, but one participant's response in the interview questions after she evaluated the arguments provides suggestive evidence that this may have been the case. Like most of the participants, when she read argument 5, she indicated that she found it unconvincing.

P27: Are you convinced by this argument? I'm going to go with a 3, because it shows only up to 26, and I feel like if you're going to use specific examples, it doesn't prove the claim is true for every single integer.

I: So there could be one...

P27: I mean it's kind of like an assumption in the proof, so because from 4 to 26 there's at least one set of primes that add up to that even integer... it's kind of like an assumption that yeah, ok, it happened for this case, so it's going to happen later on, and that might not be true by this proof.

Later in the interview, she commented that argument 5 would have been the way that she would try to prove the conjecture prior to her transition-to-proof course.

P27: I probably, in the beginning, when I was taking [the transition-to-proof course], I would have proved it like that, and then my professor probably would have murdered my answer. He would have said that my answer only proves from 4 to 26. Technically you

## Conviction, validity, and proof

only prove it from 4 to 26, so I mean I probably would have done that initially, but I don't think it proves it for all.

This suggests that the harsh grades she received for submitting empirical arguments led her to reconsider their validity. The extent that this is true for other students in a transition-to-proof course is an open research question. Administering the instrument in this study, or one like it, to students before and after they completed the course would be an efficient way of addressing this question. A final hypothesis for why that the results from this study differ from those of previous researchers is that the effect could plausibly be accounted for by attrition. Perhaps students who held empirical proof schemes withdrew from or failed their transition-to-proof course and hence were not eligible to take part in this study. Regardless of their cause, the results from this study suggest that mathematics majors' difficulties with proof do not stem from a propensity to be convinced by empirical arguments.

### 6. 2. The diagrammatic arguments

The evidence from this study indicates that the participants find some arguments relying on diagrams to be completely convincing and an acceptable form of mathematical proof. That students are convinced by these arguments is not problematic. Indeed, many philosophers of mathematics argue that it is possible to draw reliable conclusions from diagrams and this constitutes an acceptable way to derive mathematical knowledge (e.g., Brown, 1999; Giaquinto, 2005; Kulpa, 2009). Indeed, the fact that participants generally found argument 2 to be convincing might be regarded as a positive result as other studies suggest undergraduates may have overgeneralized the maxim "you can't prove by pictures" to mean that arguments incorporating diagrams can play no role in increasing one's confidence that a claim is correct (e.g., Inglis & Mejia-Ramos, 2009).

Most participants judged argument 2 to meet the standards of proof and 10 of the 28 participants thought it was a rigorous proof. The significance of these findings is debatable. On one hand, some mathematicians and mathematics educators agree that diagrammatic arguments do (or should) represent an acceptable form of proof, as is evidenced by Nelson's (1993) collection of diagrammatic proofs and Eisenberg and Dreyfus' (1991) contention that the status of diagrammatic arguments should be elevated from heuristic to demonstration in mathematics classrooms. On the other hand, the types of proof that students are expected to produce in their transition-to-proof courses are usually deductive and not diagrammatic. This suggests that the participants who judged diagrammatic arguments to be proofs may not have been aware of their professors' intentions in this regard.

### 6. 3. The deductive arguments

The participants' evaluation of deductive arguments was complex. For the three valid deductive arguments used in this study, most participants judged these arguments to be proofs and the majority found them to be fully convincing. It also appears that most participants did not judge the deductive arguments to be proofs solely because they were deductive or because they used mathematical symbols, as the prospective elementary teachers did in Martin and Harel's (1989) study. Participants rarely cited the form of the argument they were reading as a reason for rating the argument in terms of convincingness or being a proof and the majority of the participants found at least one of the seven deductive arguments not to be a proof due to its mathematical content.

Nonetheless, participants exhibited difficulty in detecting flaws in deductive arguments and were prone to accepting flawed deductive arguments as convincing valid

proofs. Further, many participants sometimes accepted an argument that they did not find fully convincing as a proof. The former phenomenon often was the result of the participants *failing* to exhibit a skill related to reading proofs, such as determining if the proof had an appropriate structure (i.e., being sure the assumptions and conclusions of the proof were appropriate) or identifying and verifying the implicit mathematical principle used to deduce a new assertion from previous ones. The latter phenomenon frequently occurred because participants harbored doubts about a particular aspect of a proof, but did not take the time to resolve these doubts.

### 6. 4. The relationship between conviction and validity

This study replicates Segal's (2000) finding that some undergraduates do not view a convincing argument and a mathematical proof to be synonymous. The interview data suggest two reasons that can account for Segal's finding. First, some participants considered different factors when deciding if an argument was personally convincing or would be sanctioned as a mathematical proof. When considering the latter judgment, some participants would cite what they believed were the sociomathematical norms with regard to proof based on comments that their professors made or common characteristics of proofs that they had observed in their mathematics courses. For instance, five participants did not believe that argument 6 met the standards of proof because it contained a graph, even though they found the graph present in the argument to be personally convincing. Second, participants were prepared to judge an argument as being a proof, even if they still held some doubts about its validity, leading them to claim an argument was a proof yet still not be fully convinced by it.



## References

- Alcock, L. & Weber, K. (2005). Proof validation in real analysis: Inferring and evaluating warrants. *Journal of Mathematical Behavior*, 24(2), 125-134.
- Balacheff, N. (1987). Processus de preuves et situations de validation. *Educational Studies in Mathematics*, 18(2), 147-176.
- Balacheff, N. (1988). Aspects of proof in pupil's practice of school mathematics. In D. Pimm (ed.) *Mathematics, teaching, and children*. London: Hodder & Stoughton.
- Chazan, D. (1993). High school geometry students' justifications for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24, 359-387.
- de Villiers (1990). The role and function of proof in mathematics. *Pythagoras*, 24, 17-24.
- Dreyfus, T. (1999). Why Johnny can't prove. *Educational Studies in Mathematics*, 38, 85-109.
- Eisenberg, T. & Dreyfus, T. (1991). On the reluctance to visualize in mathematics. In W. Zimmerman & S. Cunningham (eds.) *Visualization in teaching and learning mathematics*. Washington, D.C.: The Mathematical Association of America.
- Ericsson, K.A. & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data* (2<sup>nd</sup> ed.). Cambridge, MA.: Bradford Books/MIT Press,
- Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics*, 3(2), 9-18.
- Giaquinto, M. (2007). *Visual thinking in mathematics: An epistemological study*. Oxford University Press: Oxford.
- Hanna, G. (1991). Mathematical proof. In D. Tall (ed.) *Advanced Mathematical Thinking*. Dordrecht: Kluwer.
- Harel, G. (2001). The development of mathematical induction as a proof scheme: A model for DNR-based instruction. In S. Campbell & R Zazkis (eds.) *Learning and teaching number theory*. Ablex Publishing Company: New Jersey.
- Harel, G. & Sowder, L. (1998). Students proof schemes. *Research in collegiate mathematics education*, 3, 234-282.
- Harel, G. & Sowder, L. (2007). Towards a comprehensive perspective on proof. In F. Lester (ed.) *Second handbook of research on mathematical teaching and learning*. NCTM: Washington, DC.

- Hazzan, O. & Zazkis, R. (2003). Mimicry of proofs with computers: The case of linear algebra. *International Journal of Mathematics Education in Science and Technology*, 34, 385-402.
- Healy, L. & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31, 396-428.
- Hemmi, K. (2006). *Approaching proof in a community of mathematical practice*. Doctoral dissertation. Stockholm University. Available for download at: <http://fou.skolporten.com/art.aspx?id=a0A20000000DOi7EAG&typ=art>. Last downloaded: March 1, 2010.
- Herbst, P. (2004). Interactions with diagrams and the making of reasoned conjectures in geometry. *ZDM*, 36(5), 129-139.
- Herbst, P. & Brach, C. (2006). Proving and doing proofs in high school geometry: What is it that is going on for the students. *Cognition and Instruction*, 24, 73-122.
- Hersh, R. (1997). *What is mathematics, really?* Oxford: Oxford University Press.
- Inglis, M. & Mejia-Ramos, J.P. (2009a). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction*, 27, 25-50.
- Inglis, M. & Mejia-Ramos, J.P. (2009b). On the persuasiveness of visual arguments in mathematics. *Foundations of Science*, 14, 97-110.
- Knuth, E. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education*, 33(5), 379-405.
- Kulpa, Z. (2009). Main problems of diagrammatic reasoning: Part 1. The generalization problem. To appear in *Foundations of Science*.
- Mamona-Downs, J. & Downs, M. (2005). The identity of problem solving. *Journal of Mathematical Behavior*, 24(3/4), 385-401.
- Martin, W. G. & Harel, G. (1989). Proof frames of pre-service elementary teachers. *Journal for Research in Mathematics Education*, 20(1), 41-51.
- Mason, J., Burton, L., & Stacey, K. (1981). *Thinking mathematically*. London: Addison Wesley.
- Moore, R.C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27, 249-266.

- National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.
- Nelsen, R.B. (1993). *Proofs without words*. MAA: Washington, DC.
- Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics*, 66, 23-41.
- RAND Mathematics Study Panel (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education*. Santa Monica, CA: RAND corporation.
- Raman, M. (2002). *Proof and justification in collegiate calculus*. Unpublished doctoral dissertation. University of California. Berkeley.
- Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, 7, 5-41.
- Recio, A.M. & Godino, J.D. (2001). Institutional and personal meanings of proof. *Educational Studies in Mathematics*, 48(1), 83-99.
- Reid, D. & Inglis, M. (2005). Talking about logic. *For the Learning of Mathematics*, 25(2), 24-25.
- Schoenfeld, A. (1985). *Mathematical problem solving*. Orlando: Academic Press.
- Segal, J. (2000). Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior*, 18, 191-210.
- 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.
- Selden, J. & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics*, 29(2), 123-151.
- Simon, M.A. (1996). Beyond inductive and deductive reasoning: The search for a sense of knowing. *Educational Studies in Mathematics*, 30, 197-210.
- Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. London: SAGE.
- Stylianides, A. & Stylianides, G. (2008). Enhancing undergraduates students' understanding of proof. In *Proceedings of the 11<sup>th</sup> Conference for Research in Undergraduate Mathematics Education*. Last downloaded March 18, 2009: [http://www.rume.org/crume2008/Stylianides\\_LONG\(21\).pdf](http://www.rume.org/crume2008/Stylianides_LONG(21).pdf).

- Stylianides, A. & Stylianides, G. (2009). Proof construction and evaluation. *Educational Studies in Mathematics*, 72, 237-253.
- Tall, D. (1989). The nature of mathematical proof. *Mathematics Teaching*, 128, 28-32.
- Vinner, S. (1997). The pseudo-conceptual and pseudo-analytic thought processes in mathematical learning. *Educational Studies in Mathematics*, 34, 97-129.
- Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, 48(1), 101-119.
- Weber, K. (2004). Traditional instruction in advanced mathematics: A case study of one professor's lectures and proofs in an introductory real analysis course. *Journal of Mathematical Behavior*, 23, 115-133.
- Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, 39, 431-459.
- Weber, K. & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, 56, 209-234.
- Weber, K. & Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics*, 25(1), 34-38.

## Conviction, validity, and proof

### Appendix- Arguments used in this study

Argument 1:

*Claim:*  $(a + b)^2 = a^2 + 2ab + b^2$

*Argument:*

$$(a + b)^2 = (a + b)(a + b)$$

$$(a + b)(a + b) = a(a + b) + b(a + b)$$

$$a(a + b) = a^2 + ab$$

$$b(a + b) = ba + b^2$$

$$\text{So } (a + b)(a + b) = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2$$

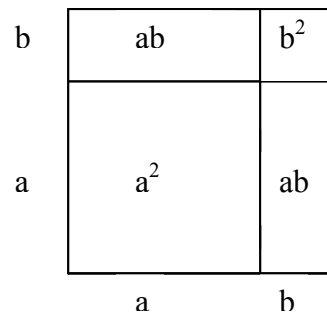
---

Argument 2:

*Claim:*  $(a + b)^2 = a^2 + 2ab + b^2$

*Argument:*

Consider the diagram below:



The length and width of the square are each  $(a+b)$ , so the area of the diagram is  $(a+b)(a+b) = (a+b)^2$ .

The area can also be found by adding the areas of the four cells of the square whose areas are  $a^2$ ,  $ab$ ,  $ab$ , and  $b^2$ , which is  $a^2 + 2ab + b^2$ .

$$\text{So } (a+b)^2 = a^2 + 2ab + b^2.$$

---

Argument 3:

*Claim:* For all natural numbers  $n$ ,  $n^3 - n$  is divisible by 6.

*Argument.*

$$n^3 - n = n(n^2 - 1) = n(n+1)(n-1).$$

Either  $n$  is even or  $n+1$  is even.

Since both numbers are factors of  $n^3 - n$ ,  $n^3 - n$  is even.

Because  $n-1$ ,  $n$ , and  $n+1$  are three consecutive numbers, one of them is divisible by 3.

So  $n(n+1)(n-1) = n^3 - n$  is divisible by 3.

Since  $n^3 - n$  is even and divisible by 3,  $n^3 - n$  is divisible by 6.

---

Argument 4:

*Claim.* The equation,  $4x^3 - x^4 = 30$ , has no real solutions.

*Argument.* Consider the function,  $f(x) = 4x^3 - x^4$ . Because  $f(x)$  is a polynomial of degree 4 and the coefficient of  $x^4$  is negative,  $f(x)$  is continuous and will approach  $-\infty$  as  $x$  approaches  $\infty$  or  $-\infty$ . Hence,  $f(x)$  must have a global maximum. The global maximum will be a critical point.  $f'(x) = 12x^2 - 4x^3$ . If  $f'(x) = 0$ , then  $x = 0$  or  $x = 3$ .  $f(0) = 0$ .  $f(3) = 27$ . Since  $f(3)$  is the greatest  $y$ -value of  $f$ 's critical points, the global maximum of  $f(x) = 27$ . Therefore  $f(x) \neq 30$  for any real number  $x$ .  $4x^3 - x^4 = 30$  has no real solutions.

---

Argument 5:

*Claim.* Any even integer greater than two can be written as the sum of two primes.

*Argument.* Consider the following table:

Even	Sum of two primes
4	2+2
6	3+3
8	3+5
10	3+7, 5+5

## Conviction, validity, and proof

12	5+7		
14	3+11, 7+7		
16	3+13, 5+11		
18	5+13, 7+11		
20	3+17, 7+13		
22	3+19, 5+17, 11+11		
24	5+19, 7+17, 11+13		
26	3+23, 7+19, 13+13		

First, note that each even number between 4 and 26 can be written as the sum of two primes. Second, note that the number of pairs of primes that work appears to be increasing. For 4, 6, 8, and 12, there is only one prime pair whose sum is that number. For 22, 24, and 26, there are three prime pairs whose sum is that number. Every even number greater than 2 will have at least one prime pair whose sum is that number. For large even numbers, there will be many prime pairs that satisfy this property.

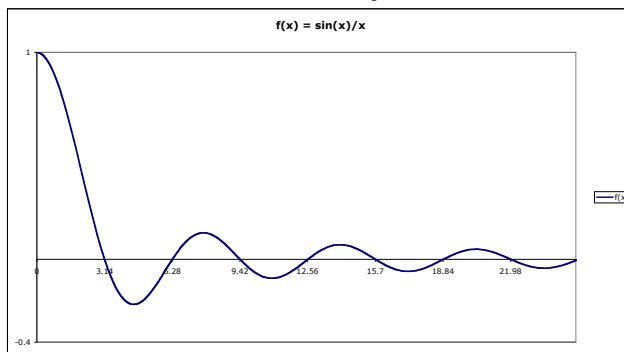
Argument 6:

Claim.  $\int_0^{\infty} \frac{1}{x} \sin x dx > 0$

Argument. The graph of  $f(x) = \frac{1}{x} \sin x$  is given below.

$\int_0^{\infty} \frac{1}{x} \sin x dx > 0$  means that  $f(x) = \frac{1}{x} \sin x$  has more area above the  $x$ -axis than below it.

To show this, note that it is clear from the graph that the first positive region—between 0 and  $\pi$  (about 3.14)—has more area than the first negative region—between  $\pi$  and  $2\pi$  (between 3.14 and 6.28). The second positive region has more area than the second negative region. The third positive region has more area than the third negative region. Since each positive region has a greater area than the negative region to the right of it, the overall area of  $\int_0^{\infty} \frac{1}{x} \sin x dx$  will be positive.



Argument 7:

Claim: If  $n^2$  is divisible by 3, then  $n$  is divisible by 3.

Argument.

We need to show that  $n$  is divisible by 3.

If  $n$  is divisible by 3, then there exists an integer  $k$  such that  $n = 3k$ .

$$n^2 = (3k)^2 = 9k^2.$$

So  $n^2$  is divisible by 9.

All numbers divisible by 9 are also divisible by 3.

So if  $n^2$  is divisible by 3, then  $n$  is divisible by 3.

Argument 8:

Claim: Let  $f(x)$  be a real valued function,  $a$  and  $b$  be real numbers, and  $b > a$ .

$$\int_a^b |f(x)| dx \geq \int_a^b f(x) dx$$

## Conviction, validity, and proof

*Argument.* (Proof by cases).

Either  $f(x) \geq 0$  or  $f(x) < 0$ .

Case 1:  $f(x) \geq 0$ .

If  $f(x) \geq 0$ , then  $|f(x)| = f(x)$ .

$$\text{Thus, } \int_a^b |f(x)| dx = \int_a^b f(x) dx.$$

Case 2:  $f(x) < 0$ .

$$\text{If } f(x) < 0, \text{ then } \int_a^b f(x) dx \leq 0.$$

$$\text{Since } |f(x)| > 0, \text{ then } \int_a^b |f(x)| dx \geq 0.$$

$$\text{So } \int_a^b |f(x)| dx \geq 0 \geq \int_a^b f(x) dx.$$

$$\text{Thus, } \int_a^b |f(x)| dx \geq \int_a^b f(x) dx.$$

---

Argument 9:

*Claim.* Let  $f(x) = \ln x$ . Then  $f(x) \rightarrow \infty$  as  $x \rightarrow \infty$ .

*Argument.*

Let  $a$  and  $b$  be positive real numbers with  $a > b$ .

Dividing both sides by  $b$  gives:

$a/b > 1$  (since  $b$  is positive).

$\ln(a/b) > 0$  (since  $\ln x > 0$  when  $x > 1$ )

$\ln(a) - \ln(b) > 0$  (by the rules of logarithms)

$\ln(a) > \ln(b)$

Hence, for positive reals  $a$  and  $b$ , if  $a > b$ , then  $f(a) > f(b)$ .

Therefore,  $f(x) \rightarrow \infty$  as  $x \rightarrow \infty$ .

---

Argument 10:

*Claim.* For all real numbers  $x$ ,  $x^2 + 12x + 28 > 0$ .

*Argument:*

$$x^2 + 12x + 28 =$$

$$(x^2 + 12x + 24) + 4 =$$

$$(x + 6)^2 + 4$$

Since  $(x + 6)^2$  is a perfect square,  $(x + 6)^2 \geq 0$  for all real numbers  $x$ .

Hence  $(x + 6)^2 + 4 \geq 4 > 0$ .

So  $x^2 + 12x + 28 > 0$ .

## Conviction, validity, and proof

Table 1

Argument Number	Reasoning used in argument	Format of argument	Mathematical Domain	Flaw in argument
1	Deductive	Symbolic	H.S. Algebra	None
2	Diagrammatic	Diagram, Prose	H.S. Algebra	None
3 (adapted from Fischbein & Kedem (1982))	Deductive	Symbolic	Number Theory	None
4 (adapted from Selden, Selden, & Mason (1989))	Deductive	Symbolic, Prose	Calculus	None
5	Empirical	Table, Prose	Number Theory	Empirical reasoning was the sole form of justification
6	Diagrammatic	Diagram, Prose	Calculus	Only displays a limited portion of the graph. Does not explain why the region areas are decreasing.
7 (adapted from Selden & Selden (2003))	Deductive	Symbolic, Prose	Number Theory	The argument begins by assuming the conclusion
8	Deductive	Symbolic, Line-by-line	Calculus	The cases described in the first line are not exhaustive
9 (adapted from Alcock & Weber (2005))	Deductive	Symbolic, Line-by-line	Pre-calculus	The last line of the argument does not follow from previous assertions
10	Deductive	Symbolic, Line-	H.S. Algebra	The third line contains an algebraic error

## Conviction, validity, and proof

Table 2

	1	2	3	4	5	6	7	8	9	10
Understanding										
5	23	26	22	14	25	22	27	22	20	28
4	2	1	4	13	2	4	1	3	4	0
3 or less	3	1	2	1	1	2	0	3	4	0
Conviction										
5	22	22	16	16	1	11	14	15	20	12
4	2	4	5	10	5	7	2	3	4	2
3 or less	4	2	7	2	22	10	12	9*	4	16
Rigorous proof	15	10	17	23	0	3	12	15	11	11
Non-rigorous	9	14	7	4	2	9	4	2	8	1
Not a proof	0	4	4	0	26	14	12	9	5	16
Unsure	4	0	0	1	0	2	0	1	2	0
Other	0	0	0	0	0	0	0	1	2	0

\*- One student declined to rate argument 8 for conviction

## Conviction, validity, and proof

Table 3

	Valid Deductive Proofs (1, 3, 4)	Invalid Deductive Proofs(7, 8, 9, 10)	Empirical Argument(5)	Diagrammatic Arguments (2,6)
Proof	75 (95%)	64 (60%)	2 (7%)	36 (67%)
Not a proof	4 (5%)	42 (40%)	26 (93%)	18 (33%)
Convincing	54 (64%)	61 (53%)	1 (4%)	33 (59%)
Not convincing	30 (36%)	54 (47%)	27 (96%)	23 (41%)

A response was coded as a *proof* if the participant evaluated the argument as a rigorous or non-rigorous proofs and *not a proof* if the participant evaluated the argument as not meeting the standards of a proof. Additional responses were not included in this tally. A response was coded as *convincing* (short for “completely convincing”) if the participant evaluated an argument with a 5 in terms of conviction and *not convincing* (short for “not completely convincing”) if the participant evaluated an argument with a 4 or less in terms of conviction. This is because a score of 5 was associated with the judgment “I was completely convinced by this argument”.

## Conviction, validity, and proof

**Table 4.**

	Problem										
	1	2	3	4	5	6	7	8	9	10	total
CP	22	22	13	16	0	6	14	15	19	11	139
NP	2	2	11	11	2	6	2	2	0	1	39
CN	0	0	3	0	1	5	0	0	0	1	10
NN	0	4	1	0	25	9	12	9	5	15	80
Other	4	0	0	1	0	2	0	2	4	0	13

CP stands for *a completely convincing proof* representing instances that were coded as a 5 for conviction and as a rigorous or non-rigorous proof.

NP stands for *non-convincing proof* representing instances that were coded as a 4 or less for conviction and as a rigorous or non-rigorous proof.

CN stands for *a completely convincing non-proof* representing instances that were coded as a 5 for conviction but not meeting the standards of a proof.

NN stands for *non-convincing non-proof* representing instances that were coded as a 4 or less for conviction and not meeting the standards of a proof.

Other represents the instance when a participant refused to rate how convincing an argument was and the 12 instances where a participant selected as “unsure” or other” when judging if the argument constituted a proof. A score of 5 was used as the threshold for conviction because a score of 5 was associated with the judgment “I was completely convinced by this argument”.