
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

The effect of authority on the persuasiveness of mathematical arguments

PLEASE CITE THE PUBLISHED VERSION

<http://dx.doi.org/10.1080/07370000802584513>

PUBLISHER

Routledge (© Taylor and Francis)

VERSION

AM (Accepted Manuscript)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Inglis, Matthew, and Juan P. Mejia-Ramos. 2019. "The Effect of Authority on the Persuasiveness of Mathematical Arguments". figshare. <https://hdl.handle.net/2134/8571>.

Running head: AUTHORITY AND MATHEMATICAL ARGUMENTS

The Effect of Authority on the Persuasiveness of Mathematical Arguments

Matthew Inglis

Loughborough University

Juan Pablo Mejia-Ramos

University of Warwick

Address correspondence to Matthew Inglis, Mathematics Education Centre, Loughborough University, Loughborough, LE11 3TU, United Kingdom. Email: m.j.inglis@lboro.ac.uk

Abstract

Three experiments are reported which investigate the extent to which an authority figure influences the level of persuasion undergraduate students and research-active mathematicians invest in mathematical arguments. We demonstrate that, in some situations, both students and researchers rate arguments as being more persuasive when they are associated with an expert mathematician than when the author is anonymous. We develop a model which accounts for these data by suggesting that, for both students and researchers, an authority figure only plays a role when there is already some uncertainty about the argument's mathematical status.

Implications for pedagogy, and for future research, are discussed.

The Effect of Authority on the Persuasiveness of Mathematical Arguments

It is widely argued that mathematical argumentation and proof should be central to the practice of learning mathematics at both the school and university levels (e.g. Hanna, 2007; National Council of Teachers of Mathematics [NCTM], 2000; Quality Assurance Agency, 2002; Schoenfeld, 1994). Consequently, students' conceptions of argumentation and proof have been extensively studied by education researchers. Harel and Sowder (1998), for example, introduced the influential *proof schemes* framework, a classification of how students gain conviction in mathematical statements. Noting that many different proof schemes have been observed in use by students, Harel emphasized that students should be encouraged to evaluate mathematical arguments by using similar criteria to those deployed by professional research-active mathematicians: "The goal of instruction must be unambiguous; namely, to gradually refine current students' proof schemes toward the proof scheme shared and practiced by the mathematicians of today." (Harel, 2001, p.188).

Other researchers have shared this view, emphasizing that expert mathematical practice is relevant for the design of pedagogy at both the school and university levels (Brousseau, 1997; Stylianou, 2002). As a consequence of similar views to this, Harel, Selden and Selden (2006) called for further research on expert and novice mathematical behavior, arguing that such research would lead to pedagogical insights.

The goal of this paper is to explore the behavior of expert mathematicians and advanced mathematics students with respect to one specific factor that may influence the level of persuasion that they invest in a mathematical argument: the authority of the argument's author. We first situate this work within the broader educational literature on mathematical argumentation and proof.

Background

Argumentative Activities in Mathematics and Mathematics Education

There is a wide variety of activities that students and mathematicians engage in, which could be described as related to mathematical argumentation and proof. In laying a preliminary map of all mathematical activities, Giaquinto (2005) provided a useful structure within which discussions of mathematical argumentation can be situated. He suggested that, for each mathematical activity, there are three associated general activities: producing, presenting, and taking in. In terms of argumentative activities these correspond to constructing a novel argument, presenting an available argument, and reading a given argument (with the aim of either evaluating or understanding it). In turn, each of these general activities can be performed with different goals in mind, which may engender different types of behavior and therefore lead to a finer categorization of mathematical activities regarding argumentation and proof.

De Villiers (1990), following Bell's (1976) initial categorization, produced a fivefold categorization of the functions of mathematical proof: verification (concerned with the truth of a statement), explanation (providing insight into why the statement is true), systematization (the organization of various results into a coherent deductive system), discovery (the discovery/invention of new results) and communication (the transmission of mathematical knowledge). Although de Villiers's primary focus was on the functions of *proof* in mathematics, he argued that the categorization also applied to other types of mathematical argumentation, including both intuitive and empirical argumentation (see also de Villiers, 2004).¹

De Villiers's categorization suggests that each of the three general activities related to mathematical argumentation presented by Giaquinto (2005) can be performed with different

goals in mind. For instance, the construction of an argument can be motivated by the estimation of the truth of a conjecture, the explanation of a given statement, its systematization, or simply to reach new results. Similarly, someone may present an argument in order to persuade a given audience of the conclusion's truth, to provide them with insight into why it is true, or to demonstrate the argument's validity in a given system. Finally, one may read an argument with the intention of understanding it, or in order to evaluate how persuasive, explanatory, or valid it is. Research has demonstrated that student behavior can greatly vary depending upon which of de Villiers's goals they have in mind. Healy and Hoyles (2000), for example, found that students preferred empirical arguments when they focused on convincing themselves of the truth of a statement (an activity related to de Villiers's verification), but preferred algebraic arguments when asked which argument they would give to their teacher (a function more closely related to de Villiers's communication). The two activities related to mathematical argumentation that have been most widely studied by education researchers are the construction of arguments to gain conviction regarding the truth of a statement, and the reading of arguments with the intention of evaluating how persuasive they are. We now briefly review earlier work on these two important areas.

Much research has focused upon how students construct arguments in order to convince themselves of the truth or falsity of mathematical statements. To be clear, in this paper we use the term *gaining conviction* to refer to the act of forming a belief about the (probable) truth or falsity of a mathematical statement.² Clearly, one important way of gaining conviction in a statement is to construct an argument with the statement as its conclusion; and Harel and Sowder (1998) produced a categorization of the types of arguments that university students produce for this purpose. They found that empirical arguments were regularly used by students to convince

themselves, an observation which has also been made of students at other educational levels (e.g. Balacheff, 1987; Bell, 1976; Coe & Ruthven, 1994; Recio & Godino, 2001). Of particular relevance for the current study, Harel and Sowder also found that students would construct, or attempt to construct, arguments based on authority (typically by citing teacher or textbook remarks) in order to convince themselves. This issue is discussed further in the next section.

Another widely studied mathematical activity, and one of relevance for the current study, consists of students evaluating how persuaded they are by a given mathematical argument. The activity of judging how persuasive one finds an argument is an important part of mathematics, and is an explicit goal of many school curricula. The NCTM's (2000) *Principles and Standards for School Mathematics*, for example, stated that "students should develop high standards for accepting explanations" and should "formulate and critique explanations so [that] their classrooms become communities of inquiry" (p. 346). In this paper we use the term *being persuaded* to refer to the act of using a given mathematical argument to gain conviction in the (probable) truth or falsity of the statement that is the argument's conclusion. Thus being *persuaded by an argument* is different from being *convinced of a statement*, as it is primarily focused on the properties of the given argument, whereas the activity of gaining conviction is primarily related to the statement and may be influenced by many (possibly unspecified) arguments, or none at all.

The activity of argument evaluation has been studied in the literature in two main ways: (i) by investigating the processes students use when reading mathematical arguments, and (ii) by investigating the types of arguments that they find persuasive.

Rav (1999) noted that the process of reading a mathematical argument is a complex one, and suggested that it typically involves extra deductive reasoning to fill any gaps left in the

argument. Such gaps are inevitable given the rarity and impractical length of strict formal derivations in mathematical texts (Fallis, 2003). Selden and Selden (2003) found that few undergraduate students engaged with such deductive gap-filling, but instead tended to concentrate on surface features when reading arguments.

Weber (2008) analyzed the ways in which research-active mathematicians read purported proofs, concentrating on the methods by which they determine whether an argument constitutes a valid proof. He found that the mathematicians in his sample did regularly gap-fill when reading incomplete arguments, but that they did not use exclusively deductive methods to do so. Instead there were many instances of informal argument construction and, perhaps more surprisingly, the evaluation of one or more examples during the mathematicians' decision making processes. Several of the mathematicians, when reflecting on their proof reading strategies, said that they might use different methods when reading student-generated arguments to arguments written by authority figures; i.e. that it was possible that they might spend longer checking that they were correct before rejecting a step in an expert's proof, compared to a student-generated proof.

As well as looking at the reading of arguments, researchers have also focused on the characteristics of arguments that students find persuasive. In a longitudinal study, Segal (1999) found that first year undergraduate students consistently rated empirical arguments, including one which consisted of the evaluation of a single example, as being highly persuasive. Knuth (2002) asked secondary school teachers to compare the persuasiveness of different arguments and found that teachers gave the highest ratings to arguments which used specific examples or visual representations, and to those whose terms and methods they were familiar with. Similar findings have been reported at the undergraduate level (Reid & Roberts, 2004).

Yackel and Cobb (1996) investigated the bases of students' beliefs about what constitutes an appropriate argument, referring to such normative aspects of classroom interaction specific to mathematics as *sociomathematical norms*. Taking a sociocultural perspective, they argued that such understandings are formed by participation in mathematics classroom cultures, and are the result of the interaction and implicit negotiation between students and teachers. Similar kinds of norms have also been found to play a significant role in the context of an undergraduate education. Yackel, Rasmussen, and King (2000), for example, found that within the classroom culture of a differential equations course “because that’s the rule” constituted a legitimate mathematical explanation; in contrast entirely procedural explanations were considered illegitimate, and were challenged.

Although many studies have focused on the factors (and their causes) which affect how persuasive school and university mathematics students find arguments, we are aware of no previous research related to how research-active mathematicians make such judgments. This is particularly surprising if—as Brousseau (1997), Harel (2001) and others have argued—pedagogical practice related to proof should take as its goal the development of students' behavior so that it better matches the practice of mathematicians.

In the current paper we add to the growing literature on the factors which influence the persuasiveness of arguments by exploring one factor in detail: the presence or absence of an expert author. First we briefly review earlier work which has discussed this and related issues, with reference to both conviction and persuasion.

The Role of Authority in Everyday and Mathematical Argumentation

Appeals to expert opinion are widespread in everyday situations. Indeed, Walton (1997) went as far as to say that “nearly everything we believe is believable because it is based on the

opinion of experts” (p.1). Consequently, it is no surprise that arguments which are from highly credible sources are seen as being more persuasive than those which are not (e.g. Reinard, 1988; Sternthal, Phillips, & Dholakia, 1978). Walton (2006), in a survey of various types of defeasible argumentation, explicated the appeal to expert opinion (or *argumentum ad verecundian*) by laying out its argumentation scheme:

Major Premise: Source E is an expert in subject domain D containing proposition P .

Minor Premise: E asserts that P (in domain D) is true (false).

Conclusion: P may plausibly be taken to be true (false).

In the context of mathematics, two subtly different types of propositions are relevant for the current discussion. The first is where P is a straightforward mathematical statement, an example of which might be a theorem or lemma. In our terms as outlined above, this type of appeal to expert opinion relates to *conviction*, as P is a statement. A different type of appeal to expert opinion, which focuses upon the *persuasiveness* of an argument, can be made. This type has the following argumentation scheme:

Major Premise: Source E is an expert in subject domain D containing argument A .

Minor Premise: E asserts that argument A (in domain D) justifies (to a given extent) proposition P .

Conclusion: A may plausibly be taken to justify (to a given extent) P .

The distinction between conviction and persuasion leads to two different types of appeals to expert authority; and both types have been the subject of discussions in the literature.

In an in-depth analysis of the role of proof in mathematics and mathematics education, Hanna (1983) asked what factors influence whether a theorem becomes accepted in the mathematical community. Apparently accepting that appeals to expert authority of the first type

do play an important role in mathematical practice, Hanna (1983) asserted that whether or not “the author has an unimpeachable reputation as an expert” is a factor which ranks above rigorous proof in determining whether a theorem will gain acceptance among mathematicians (p. 70). Reid (2002) enthusiastically endorsed Hanna’s analysis and went further, speculating that if professional mathematicians accept appeals to expert authority, then it could be reasonable to describe structurally similar arguments authored by students as being ‘mathematical’ and therefore legitimate.

However, many researchers have apparently disagreed with Hanna’s (1983) and Reid’s (2002) claims about expert mathematical practice. In their comprehensive review of the types of evidence that students use to convince themselves of the truth of mathematical statements, Harel and Sowder (1998) found many instances of the “authoritarian proof scheme,” which essentially refers to appeals to expert opinion. Based on the view that the goal of instruction is to develop students’ proof schemes so that they match those shared by modern mathematicians, Harel and Sowder (2005) described the authoritarian proof scheme as “an undesirable, yet common, way of thinking” and suggested that instruction must “institute a didactical contract that attempts to suppress [it]” (Harel & Sowder, 2007). Cobb and Yackel (1996) made a similar suggestion, by citing the development of the sociomathematical norm of “intellectual autonomy”—where reliance on an external authority is not considered acceptable—as a “central pragmatic goal” of their work with mathematics classroom cultures (p. 179). A similar point was made by Stylianides (2007) in his discussion of how the teaching of proof in early educational settings could become “intellectually honest” to the discipline of mathematics. He gave the example of an appeal to expert opinion (of the first type) given by a student and approvingly noted that the classroom teacher disqualified the argument and instead “pushed the students to come up with a

mathematical argument” (p. 309, emphasis in the original). In summary, both Harel and Sowder and Stylianides expressed a belief that (i) students’ arguments should be primarily valued in terms of whether they match the kinds of arguments that form part of expert mathematical practice; (ii) students should be discouraged from making appeals to expert authority; and (iii) instruction should attempt to suppress such appeals. This line of argument appears to be based on the belief that, contrary to Hanna’s (1983) and Reid’s (2002) views, mathematicians do not make appeals to expert authority in their mathematical work.

Hanna (2007) also discussed appeals to expert opinion with reference to the persuasiveness of purported proofs:

It is in the very nature of proof that the validity of the conclusion flows from the proof itself, not from any external authority. Proof conveys to students the message that they can reason for themselves, that they do not need to defer to authority. (p. 31).

Note that Hanna (2007) was not contradicting her earlier work. In this later discussion, she referred to appeals to expert opinion of the second type: where the proposition involved is not a mathematical statement (the case discussed by Hanna, 1983), but a claim about the persuasiveness of a purported proof. Other researchers have agreed on this point: Selden and Selden (2003), for example, wrote that “like midcentury structural critics, mathematicians seem to treat a proof as being independent from its author” (p. 6). Again, according to these researchers, mathematicians do not use appeals to expert opinion when evaluating the persuasiveness of purported proofs. However, as discussed above, Weber (2008) noted that some mathematicians claimed that they might use different methods to evaluate a student-generated proof compared to a proof written by a mathematician. In her discussion of Perelman’s purported proof of the Poincaré Conjecture, Jackson (2006) reported one mathematician’s views that

Perelman's proof "must be right" because (i) if it were not, the collective expertise of the mathematical community would have found the mistake; and (ii) Perelman's earlier work had been reliable (p. 899). Both of these statements seem to constitute appeals to expert opinion about the persuasiveness of Perelman's proof.

To summarize these earlier discussions, some researchers have argued that appeals to expert opinion are used by mathematicians to gain conviction in statements (Hanna, 1983; Reid, 2002), but not to assess the persuasiveness of arguments (Hanna, 2007). Others have argued that mathematicians may be influenced by such appeals when engaged in this latter activity (Weber, 2008); and some have implied that experts do not take into account authority figures when gaining conviction in statements (Harel & Sowder, 2005; Stylianides, 2007). Although it is clear that appeals to expert opinion do play an important role in everyday argumentation, it appears that there is little consensus in the literature as to the status of such appeals in expert mathematical practice.

All the discussions about expert mathematical practice reviewed in this section were based upon philosophical/historical analyses, introspective reflections reported by mathematicians, introspections of the authors' own practices, or impressions of the general mathematical culture. Crucially, none appears to be based on empirical studies of actual mathematical behavior. Our goal in this study was to add to these discussions by providing relevant empirical evidence; specifically, by asking: (i) Does the presence of an expert author increase the persuasiveness of mathematical arguments? (ii) Does this factor depend upon the type of argument? and (iii) Are research-active mathematicians and undergraduate students similarly affected by the presence of an expert author?

Experiment 1

The goal of Experiment 1 was to directly compare the persuasiveness of various different types of mathematical arguments in two conditions: when the readers knew that the argument was written by an authority figure, and when they did not. If knowledge of an expert author has no effect upon participants' judgments about the persuasiveness of arguments, one would expect no significant differences between the mean reported persuasion ratings of the two conditions.

Method

We used the internet to present our tasks and gather our data in order to maximize our sample size. As noted by Reips (2000), this method does present some practical difficulties, in particular the problem of multiple submissions from the same individual. We adopted the strategy advised by Reips, and implemented by Johnson-Laird and Savary (1999), and logged the IP address of each participant and the time they submitted their responses. Under the assumption that each IP address was associated with a unique individual, these data were used to screen for possible cases of multiple submission.

Web-based experimental methods have been found to produce results that are consistent with those found by traditional methods (Gosling, Vazire, Srivastava, & John, 2004; Krantz & Dalal, 2000), rendering it difficult to suggest that internet data are less valid than traditional data. Given these findings, our adherence to Reips's (2000) guidelines, and the impracticality of obtaining large samples of research-active mathematicians in any other fashion, we believe the use of a web-based approach is justified in this instance.

Participants. Participants were 384 unpaid volunteers. A total of 190 research-active mathematicians employed at Australian and British universities took part, together with 194 second and third year undergraduates from three highly ranked British universities. All three

universities only recruited students who had been very high achievers during their school studies,³ and had syllabi that emphasized formal rigorous proof from the outset of the course.

Each participant was recruited by means of an email message, sent from their departmental secretary, which explained the task and asked them to visit the experimental website should they wish to participate. The researchers and students participated through identical websites hosted at different addresses (in order to ensure that the two groups' responses were kept apart), and before each researcher commenced the task they were asked to declare that they were a "research-active mathematician."

Materials. Three different arguments were used: a heuristic argument about the digits of π , an induction argument about the size of the n -th prime, and a visual argument about the fixed point theorem. The three arguments were selected to be challenging enough so as not to be trivial for research-active mathematicians, but also not so complicated that undergraduates could not understand them. Furthermore, each of the arguments needed to have been written by an authoritative source. The heuristic and visual arguments were based on published work by two highly respected mathematicians: Professors W. T. Gowers and J. E. Littlewood respectively (Gowers, 2006, p. 194; Littlewood, 1953, p. 37).⁴ The induction argument was taken from an undergraduate textbook (Jones & Jones, 1998, p.26). The three arguments are given in full in the Appendix.

Procedure. Once a participant had loaded the experimental website, they were presented with one of the three arguments and randomly assigned into one of two conditions: those who were told the identity of the author (the named condition), and those who were not (the anonymous condition). The named and anonymous conditions were identical, except for the

presence or absence of the identity of the author (see the Appendix for details of how this information was phrased).

Participants were given the instruction:

“After having read this argument please say to what extent you are persuaded by it. If you would like to comment, or explain your selection, please do so.”

A sliding bar was used to record each participant’s reported level of persuasion in the argument, from 0% (*not persuaded*) to 100% (*totally persuaded*); and a text box was used to collect comments from those participants who wished to leave them. Upon completion of the task, participants submitted their response and the next argument was loaded. For each new argument participants were randomly reallocated into either the named or the anonymous conditions. The order in which the arguments appeared followed a counterbalanced Latin square design. As the experimental design included the anonymous control condition, and since participants were assigned to conditions randomly, any possible between-conditions differences in participants’ mean reported level of persuasion could only be attributed to the difference in the conditions: the presence or absence of an authority figure.

Results

The mean response from each group for each argument are shown in Table 1 and graphed in Figure 1.⁵ There were large individual differences in how participants from both groups responded to each argument. When responding to Gowers’s heuristic argument in the anonymous condition, for example, one researcher responded with 0% and wrote “do you really think a ‘research mathematician’ will be persuaded by this?” Another, responding to the same argument, and in the same condition, gave a rating of 100% and wrote “this argument completely persuades me that the conjecture is reasonable.” Perhaps these large individual differences can be

attributed to different interpretations of the instructions. Nevertheless, searching for between-condition differences in mean responses allows us to interrogate the influence that the presence of an authority figure had on participants' reported level of persuasion in the arguments.

Insert Table 1 about here

Insert Figure 1 about here

A univariate analysis of variance (ANOVA) with three factors (group, argument and condition) was conducted. A significant three-way interaction was found, $F(2,1094)=4.35$, $MSE=980.7$, $p=0.013$, suggesting that separate analyses should be conducted for each argument.

Students reported higher levels of persuasion in the heuristic argument than the researchers, $F(1,370)=5.81$, $MSE=928.5$, $p=0.016$. Crucially, there was also a significant main effect for condition, $F(1,370)=5.45$, $MSE=928.5$, $p=0.020$, showing that those participants who knew the argument was written by Gowers rated the argument as more persuasive than those who did not (a mean difference of 8%). When the analysis was restricted to only the researchers, this effect retained significance, $t(178)=2.19$, $p=0.030$. There was no significant group \times condition interaction effect, $F<1$.

Surprisingly, for the visual argument the group \times condition interaction effect was significant, $F(1,372)=12.2$, $MSE=1095.0$, $p=0.001$. Considering each group separately revealed that the researchers reported a higher level of persuasion if they knew it was written by Littlewood, with a mean difference of 17%, $t(188)=3.46$, $p=0.001$; but this was not the case for

the students, who showed a small and non-significant trend in the opposite direction $t(184)=1.41$, NS.

The researchers found the induction argument to be more persuasive than the students, $F(1,352)=4.40$, $MSE=914.8$, $p=0.037$. However, there was no main effect for condition, nor a group \times condition interaction, $F_s<1$.

Discussion

The results from Experiment 1 suggest that, in some cases at least, both undergraduate students and research-active mathematicians are influenced by the presence of an authority figure when evaluating mathematical arguments. Participants reported a higher level of persuasion for the heuristic argument when they knew it was written by Gowers, but there was no difference for the textbook's induction argument. The data from the visual argument, however, were more complex. The researchers who knew the argument had been written by Littlewood ranked it as being more persuasive than those who did not know; this was not the case for the students. These data are discussed in detail in the next section, beginning with the results for the visual argument.

Accounting for the visual group \times condition interaction. One initial hypothesis to account for the significant group \times condition interaction effect on the visual argument might be that the students were not as familiar with Littlewood's status as an expert mathematician as the researchers. However, if this were the sole reason, one would have expected similar group \times condition interactions in the heuristic and visual arguments, as Littlewood is at least as well known as Gowers. (In fact the group \times condition \times argument interaction effect was significant for the visual and heuristic arguments, $F(1,742)=3.96$, $MSE=4002.5$, $p=0.047$, indicating that the group \times condition interactions for these two arguments were different).

To gain deeper insight into the reasons behind the visual argument group \times condition interaction, we conducted an analysis on the explanatory comments left by participants.

Approximately 20% of participants left comments. For each argument, in each condition, the difference between the mean level of persuasion reported by comment-leavers and non-comment-leavers did not approach significance, all $ps > 0.15$, suggesting that those participants who left comments were not unrepresentative of the entire sample. The two authors of this paper independently assigned each comment to one of five pre-defined categories; these codings were then compared and the few disagreements discussed and resolved. One category (about the explanatory value of the argument) was only used by two participants, and so was incorporated into the “other” category.

Insert Table 2 about here

The range of explanatory comments given by each group is shown in Table 2. The two groups had a different range, $\chi^2(3) = 20.9$, $p < 0.001$, with the researchers more often discussing how easy or difficult it would be to turn the argument into a proof, and the students more often stating that a picture is not a proof. Indeed, many student comments in this category emphasized that they had been *taught* that a picture is not a proof. One student, for example, wrote that “trying to prove with a graph is a cardinal sin, as I have been taught.” In contrast, few of the researchers took such a position, and instead tended to concentrate on factors such as how easy it would be to turn the argument into a formal proof.

These data suggest that many students were reluctant to give any credence to visual arguments. Littlewood (1953) himself discussed this issue, remarking that “my pupils *will* not

use pictures, even unofficially.” This phenomenon has also been documented by several mathematics education researchers (e.g. Dreyfus, 1994; Vinner, 1989). Littlewood put the blame on what he saw as inappropriate pedagogical strategies:

A heavy warning used to be given [by lecturers] that pictures are not rigorous; this has never had its bluff called and has permanently frightened its victims into playing for safety. Some pictures, of course, are not rigorous, but I should say most are (and I use them whenever possible myself). (p.35).

It seems reasonable to hypothesize that this reluctance on the part of students to trust visual arguments could be the result of a sociomathematical norm (in the sense of Yackel & Cobb, 1996) initiated by the kinds of pedagogical interventions described by Littlewood, and then reinforced through students’ and lecturers’ activities in lecture theatres and seminar rooms.

This factor could account for the group \times condition interaction detected for the visual argument. Many students were not influenced by the presence or absence of the authority figure as they believed, as a consequence of the sorts of pedagogical strategies mentioned by Littlewood, that visual arguments are not acceptable. In comparison, the researchers seemed to operate in accordance with a more flexible norm regarding the validity of visual arguments, and so were more open to being persuaded by them.⁶

Accounting for the induction argument non-effect. Participants in both groups and in both conditions ranked the induction argument as being highly persuasive (with mean ratings of around 80%), and there was no significant differences between the mean level of persuasion recorded in each condition. We see three possible ways of accounting for this. One hypothesis is that mathematicians (and talented undergraduates) completely ignored the source of a purported *formal* proof when judging how persuasive it was. Another possibility is that the induction

argument was too simple to reveal any effect. In more complicated cases where there is substantial doubt about a purported proof (such as the Perelman proof discussed by Jackson, 2006) perhaps there would be an effect. One final possibility would be that the authority figure for this proof—an undergraduate textbook—was not as authoritative as either Gowers or Littlewood.

In order to investigate this issue further we replicated Experiment 1 using a more complex formal proof, with a better known authority figure, in place of the induction argument.

Experiment 2

The main goals of Experiment 2 were (i) to replicate the main results in Experiment 1, (ii) to investigate the effect that an authority figure has on the level of persuasion that mathematicians invest in more complex formal proofs than the induction argument used in Experiment 1, and (iii) to investigate whether or not a more authoritative figure than an undergraduate textbook can make a formal proof more persuasive. In view of the need to use a considerably more complex proof than the induction argument, it was decided to collect data from research-active mathematicians, but not undergraduate students.

Method

Participants were 71 research-active mathematicians employed at universities in the United States.⁷ An identical procedure and set of materials were used to that in Experiment 1, with the exception that the induction argument was replaced by a more complicated argument about the area of a polygonal's tangential triangle. This latter argument was written by George Pólya (and had been adapted from the version published by Aigner and Ziegler, 2000, p. 103). It is given in full in the Appendix.

Results

Participants' mean reported levels of persuasion in the three arguments are shown in Table 3. In the heuristic and visual arguments the pattern of responses essentially replicated those found in Experiment 1. Participants in the named condition had a higher mean reported level of persuasion in the visual argument than those in the anonymous condition, $t(65)=2.93$, $p=0.003$, one-tailed.⁸ A similar result was found in the heuristic argument, $t(69)=1.78$, $p=0.040$, one-tailed. For both arguments, the magnitudes of the difference in means were approximately equal to the equivalent differences for the researcher group in Experiment 1.

Insert Table 3 about here

However, in the Pólya argument, there was no difference between the mean reported level of persuasion, $t(58)=0.003$, NS, with participants in both conditions rating the argument as highly persuasive.

Discussion

The results from the heuristic and visual arguments replicated the results from the researcher group in Experiment 1, giving confidence that these findings are real effects. However, despite the increased complexity of the Pólya argument compared to the induction argument, there was no difference in the mean reported level of persuasion for each condition. This result would appear to rule out the third hypothesis put forward to explain the lack of effect in the induction argument in Experiment 1—that the undergraduate textbook was not authoritative enough. George Pólya is a well known and highly respected mathematician, and (one would think) a mathematician who would be held in similar levels of esteem as both

Gowers and Littlewood (who were apparently sufficiently authoritative to produce effects on the heuristic and visual arguments).

However, the data from Experiment 2 are unable to distinguish between the first and second hypotheses. It may be that an effect *would* be found for extremely complex formal proofs, such as the Perelman case discussed in the introduction. Of course, Perelman's purported proof of the Poincaré conjecture is of several orders of magnitude more complex than the Pólya argument used in Experiment 2; it may be that a formal proof too complex to be practical for an empirical study would be needed to detect any effect. Alternatively, it may be that mathematicians *completely* ignore the source of an argument that looks like a formal proof.

Although the data from Experiments 1 and 2 give us confidence that the effects detected for the visual and heuristic arguments are real—the presence of an authority figure caused participants to report higher level of persuasion in an argument than they would otherwise do—it could be that the difference was an unintended consequence of the instructions that we used. Rather than actually affect a person's level of persuasion in the argument, the presence of an authority figure might merely have altered their interpretation of the question.

Although participants had been asked to rate the extent to which they were persuaded by the argument, evidence from the explanatory comments left by participants suggested that some participants had interpreted these instructions in different ways. For example, one participant responded to the heuristic argument with a rating of 2% and commented that “if a manuscript that made an analogous argument came to me for refereeing, I'd recommend it be rejected for lack of mathematical rigor.” Another gave an 80% rating and wrote, “After reading the argument, if the claim were proved to be false I'd be surprised and intrigued.” The first participant seemed to have interpreted the instructions as a question about the admissibility of the

argument in the context of an academic mathematics journal—whether the argument met the expected standards for presentation in such a context. In contrast the second participant seems to have interpreted more in the manner we intended—by evaluating the extent to which they were persuaded by the argument, i.e., the extent to which they believed the argument allowed them to gain conviction in the truth of the argument’s conclusion.

One way of accounting for the results of Experiments 1 and 2 is to suggest that the presence of an authority figure changed the way in which participants interpreted the request to rate their level of persuasion. For example, the presence of an authority figure might have biased participants away from conducting an evaluation that focused on the admissibility of the argument in a given context (as in the first example above), and towards conducting an evaluation about the persuasiveness of the argument (as in the second example above).⁹ To rule out this possibility we conducted Experiment 3.

Experiment 3

The main goal of Experiment 3 was to test whether the results of Experiment 1 and 2 survived if the question was altered in such a way as to encourage participants to evaluate the extent to which they believed the argument allowed them to gain conviction in the truth of the argument’s conclusion. If the effects found in Experiments 1 and 2 were due to the presence of an authority figure altering participants’ interpretations of the instructions, one would expect these effects to be eliminated in Experiment 3.

Method

Participants were 214 research-active mathematicians from universities affiliated with the Association of Commonwealth Universities, and from universities in the United States. An identical set of materials and procedure was used to Experiment 1, except that all the statements

being discussed were referred to as ‘claims’ (rather than theorems or conjectures), and the instructions read:

Now say to what extent you are persuaded that the claim is true, given the information, and only the information, that is contained in the argument. If you would like to comment, or explain your selection, please do so.

The aim of these alternative instructions was to bias participants towards evaluating their level of persuasion, as we have used the term in this paper; i.e. to determine the extent to which the *given argument* allows them to gain conviction in the (probable) truth or falsity of the argument’s conclusion.

Results and Discussion

The mean responses to the three arguments are shown in Table 4. In the heuristic and visual arguments, the pattern of responses essentially replicated those found in Experiments 1 and 2. Participants in the named condition found the heuristic argument to be more persuasive than those in the anonymous condition, $t(212.0)=2.75, p=0.007$, with the magnitude of the difference in means similar to that found in the previous experiments. A similar result was found in the visual argument, $t(208)=2.37, p=0.019$, although the magnitude of the difference of the means was somewhat less than those found in Experiments 1 and 2. As before, no significant between-conditions difference in ratings was found for the induction argument, $t(200.8)=1.83, NS$.

Insert Table 3 about here

The results of Experiment 3 reduce the likelihood that the effects detected in Experiments 1 and 2 were the result of the presence of an authority figure biasing participants into a different interpretation of the question. If that account were correct we would have expected the between-conditions differences to have been eliminated in Experiment 3, the instructions of which pushed participants into conducting consistent evaluations. In Experiment 3 the instructions emphasized the importance of evaluating the persuasiveness of the argument only given the information contained in the argument, and the main effects found for the research-active mathematicians in Experiments 1 and 2 were replicated.

General Discussion

Summary of Main Findings

In this paper three experiments have been reported which demonstrate that, in some situations, mathematics students and researchers report higher level of persuasion in mathematical arguments that are associated with an authority figure than in those not so associated.

Experiment 1 demonstrated that, for research-active mathematicians, authority figures increased the persuasiveness of two non-formal arguments, one heuristic and one visual. Surprisingly, although undergraduate students also reported higher levels of persuasion when they knew the identity of the heuristic argument's author, they did not exhibit the same behavior on the visual argument. Neither group was more persuaded by a formal induction proof if they knew it came from a textbook. In Experiment 2 it was shown that the reason for a lack of an effect on the induction argument was not that the proof was too simple or that the textbook lacked authority: a similar result was found using a more complex proof written by George Pólya, a highly authoritative mathematician. Finally, Experiment 3 suggested that the results of

Experiments 1 and 2 were unlikely to be artifactual consequences of the task's instructions. The task was varied slightly to encourage participants to interpret the instructions in a consistent manner, and similar effects were found to those in the previous experiments.

A Model

Experiments 1, 2, and 3 suggest that, in some cases, students and mathematicians are influenced by an authority figure when evaluating mathematical arguments. However, this was not the case for every argument. One model which accounts for these different cases is illustrated in Figure 2. The model suggests that the source of an argument only influences the reader's reported level of persuasion if, having read the argument, the reader is left uncertain about it. That is to say, an authority figure can only operate on the *uncertainty* associated with an argument.

Insert Figure 2 about here

The data regarding each of the arguments used in this paper can now be accounted for in turn. For Gowers's heuristic argument, neither the students nor the researchers seemed clear as to whether the argument was suitable to support its (non-absolute) conclusion. In terms of the model, there was uncertainty associated with the argument; and this uncertainty was operated on by the authority figure, leaving those participants in the named condition with a higher mean reported level of persuasion than those in the anonymous condition. For the induction argument the reverse was true: neither students nor researchers were left with much doubt as to the sufficiency of the argument. Consequently there was no uncertainty for the presence of the authority figure to operate on, and no effect for condition was detected. A similar explanation

accounts for the lack of an effect for researchers on the Pólya argument. Whether this would be the case for *all* formal proofs, or whether an extremely complex formal proof would result in sufficient doubt to cause an effect for authority, remains an open question.

The data from the visual argument require a more complex explanation. We suggest that the students in Experiment 1 tended to be left with *less* uncertainty about the argument's suitability than the researchers. As a consequence of a sociomathematical norm related to the status of visual arguments, many students seemed clear that the argument was not mathematical. There was no uncertainty on which the authority figure could operate, and consequently no effect for condition was found. The researchers, in contrast, appeared to have a different norm associated with visual arguments. As shown above, only two comment-leaving researchers rejected the argument simply by virtue of its visual status, but 12 instead concentrated on how close the argument was to a formal proof. In terms of the model, this more flexible view of the status of visual arguments led the researchers to have *more* doubts about the suitability of the argument. Consequently, there was more uncertainty on which the authority figure could operate, and a (large) effect was found for condition, with those knowing that Littlewood had written the argument having a higher mean level of persuasion than those who did not.

A remaining issue is to explain why an expert authority figure caused participants to increase their reported level of persuasion in arguments about which they were uncertain. One participant's explanatory comment (with reference to the heuristic argument) indicates a possibility: "We are told the argument is made by a reputable mathematician, so we implicitly assume that he would tell us if he knew of any evidence or convincing arguments to the contrary." (Researcher).

The researcher here suggested that Gowers's reputation gave confidence that he was not being deliberately misleading by hiding any evidence against his conclusion. It is also plausible (although not explicitly stated by this researcher) that Gowers's expertise as a mathematician suggested to participants that if any counter-evidence existed he would, in all likelihood, be aware of it. A similar suggestion accounts for the visual argument; researchers, being uncertain of its mathematical status, could reasonably conclude that Littlewood, being a renowned mathematician, would (i) have sufficient background knowledge to be aware of any conflicting evidence if it existed (any unusual counterexamples, for example), and (ii) be trustworthy enough to not withhold any such evidence.

The model allows for several further factors which may affect a person's level of persuasion in an argument. Although all the cases reported in this paper were of a reader's uncertainty generated by concerns over the nature of a mathematical argument, we suggest that this may not be the only factor which creates uncertainty. An argument may simply be too hard to understand, or perhaps a reader may not want to invest sufficient effort into understanding it. Although we have not tested this hypothesis, the model predicts that the presence of an authority figure would have an effect in cases of this type.

In summary, the model illustrated in Figure 2 suggests that, for both mathematics researchers and mathematics students, an authority figure *does* influence their level of persuasion in an argument, *but only when there is already some uncertainty about the argument's mathematical status*. The model accounts for all the data reported in this paper, and also makes several specific predictions suitable for future testing.

Relations to Other Disciplines

As discussed in the introduction, several researchers have designed instruction that attempts to develop students' "intellectual autonomy" by devaluing appeals to expert opinion within the mathematics classroom (e.g. Cobb & Yackel, 1996; Harel & Sowder, 1998; Stylianides, 2007). This practice contrasts sharply to other disciplines; a key goal of history education is, for example, to develop students' abilities to judge a source's reliability (e.g. Pontecorvo & Girardet, 1993). In this section we briefly contrast the findings reported in this study to related findings from an area where the influence of expert opinion has been extensively studied: the process of academic peer review.

In some disciplines single-blind peer review processes—where the reviewers are aware of the authors' identities and affiliations—are commonplace. The evidence on whether this practice creates bias is conflicting, and seems to be dependent upon the discipline involved. Evidence suggests that psychologists and economists are influenced by the names and affiliations of authors when reviewing papers (Blank, 1991; Peters & Ceci, 1982), but that medical researchers are not (van Rooyen, Godless, Evans, Smith & Black, 1998).

There are several important differences between the process of reviewing a research article and the task reported in this paper. Reviewers, as well as evaluating the extent to which they are persuaded by the argument presented in a research paper, must also evaluate, among other things, the article's significance and clarity. Nevertheless, the results of research on peer review are not inconsistent with the model we propose above. One might speculate that, in disciplines such as economics and psychology, articles more often report studies which incorporate compromises that are difficult to evaluate objectively, inevitably leading to some uncertainty in the evaluation process. In comparison, medical researchers typically have very

strict objective guidelines as to what constitutes a persuasive article (see, for example, Begg et al., 1996). This factor may lead to a greater proportion of reviewers in social science disciplines being uncertain of a manuscript's persuasiveness than in natural science disciplines, which in turn may lead to (following the model) more reviewers being influenced by the author's affiliation and name.

Implications

Earlier researchers have disagreed about the normative status of appeals to expert opinion in mathematical reasoning. Hanna (1983) and Reid (2002) both suggested that mathematicians use appeals to expert opinion *to gain conviction in mathematical statements*. In contrast, Harel and Sowder (1998, 2007) and Stylianides (2007) both suggested that using authority figures in this fashion is mathematically undesirable. It is important to emphasize that the data reported in this paper do not allow us to distinguish between these positions. Our study concentrated on whether an appeal to expert opinion affects the *persuasiveness of an argument*. Hanna (2007) and Selden and Selden (2003) suggested that the persuasiveness *of purported proofs* is evaluated independently of their authors. Our study did not produce evidence against this particular belief (indeed, we found that the presence of an authority did not affect mathematicians' evaluation of two purported proofs), but showed that when evaluating the persuasiveness of *other types of arguments* in mathematics, both mathematicians and mathematics students can be influenced by their source. Moreover, our model predicts that, contrary to the position asserted by Selden and Selden (2003), factors like extreme difficulty and time restrictions may lead mathematicians to appeal to expert opinion when evaluating the persuasiveness of purported proofs. Such factors are likely to lead to high levels of uncertainty in the written argument, and thus would increase the influence of an associated authority figure. Future research on the effect of these and other

factors will improve our understanding of the ways in which students and mathematicians read and evaluate different types of arguments in mathematics.

As a result of their views on the undesirability of appeals to expert opinion, several researchers have designed instruction with the explicit aim of discouraging students from taking into account authority figures (e.g. Cobb & Yackel, 1996; Stylianides, 2007). The data reported in this paper, however, suggest a subtly different approach. When research mathematicians are uncertain about the mathematical status of an argument, they too seem to be influenced by the argument's source. But, crucially, the presence of a source appears to have no effect when participants have little uncertainty about the argument's status. During their studies, weaker students are disproportionately likely to be uncertain about the arguments they meet, so it is not surprising that it is they who are regarded as being more likely to rely upon authority figures to judge arguments. However, we suggest that relying upon, or looking for, an authority figure should not be seen as undesirable behavior per se. Instead, the emphasis should be placed on empowering students by ensuring that they are in a position where they do not need to take into account an authority figure: where they have little uncertainty about the mathematical status of formal arguments.

Notwithstanding this suggestion, the case of the visual argument indicates that a greater mathematical maturity is not always correlated with less uncertainty over all types of argument. We have argued that, as a group, the researchers in our study were *more* uncertain about the validity of Littlewood's visual argument than the students. Because of (what Littlewood saw as) inappropriate pedagogical strategies, many undergraduate students are apparently unpersuaded by visual arguments—regardless of their source. If this interpretation is correct, the pedagogical practices of undergraduate mathematics lecturers may support, if not develop, sociomathematical

norms in the lecture hall that differ significantly from those shared by practicing mathematicians. In the particular case of the persuasiveness of visual arguments in mathematics, lecturers may be conveying only part of the story: students seem to be responding to the “heavy warning” that pictures may be misleading, while letting pass the potential of visual arguments for providing insight and suggesting results in mathematics. Exploring in more detail the range of sociomathematical norms researchers and undergraduates act in accordance with in relation to visual arguments, and whether or not the differences between these two sets of norms are desirable, would appear to be an interesting area for future research (c.f. Arcavi, 2003; Dreyfus, 1994; Vinner, 1989).

One final implication of the studies reported in this paper concerns the differences between expert and novice mathematical behavior. We agree with those researchers who have suggested that one of the major aims of an undergraduate education in mathematics should be to develop students’ skills so that they more closely match expert behavior (e.g. Brousseau, 1997; Harel et al., 2006; Stylianou, 2002). Recently, however, Weber (2008) demonstrated that expert mathematical behavior is complex and sometimes unexpected. He found that mathematicians would regularly use non-deductive methods—even including the evaluation of a single example—when bridging gaps in formal deductive proofs. Such behavior is surprising given the low status given to empirical reasoning in formal mathematical discourse, and resulting educational frameworks (e.g. Balacheff, 1987). The results of the current study support Weber’s view that expert mathematical behavior can be unexpectedly multifaceted. Indeed, many of the mathematicians who participated in the study and asked to be debriefed expressed surprise at the results. It is clear that expert mathematical behavior does not necessarily match the received view of expert mathematical behavior.

Mathematics educators have, when constructing an understanding of the nature of mathematics, typically relied upon philosophical accounts, introspective reports from mathematicians, and historical analyses of the development of mathematical concepts. While these methods are all undeniably useful for gaining a deeper understanding of the nature of mathematics, we believe that empirical research has an important contribution to make in the development, testing, and refinement of models of expert mathematical practice.

Acknowledgements

We thank Mike Sharples, Keith Weber and three anonymous reviewers for their valuable comments and suggestions.

References

- Aigner, M., & Ziegler, G. (2000). *Proofs from the book* (Second ed.). Berlin: Springer-Verlag.
- Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics*, 52, 215-241.
- Azzouni, J. (2005). Is there still a sense in which mathematics can have foundations? In G. Sica (Ed.), *Essays on the foundations of mathematics and logic* (pp. 9-47). Monza: Polimetrica.
- Balacheff, N. (1987). Processus de preuves et situations de validation. *Educational Studies in Mathematics*, 18, 147-176.
- Begg, C., Cho, M., Eastwood S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D. & Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association*, 276, 637-639.
- Bell, A. W. (1976). A study of pupils' proof conceptions in mathematical situations. *Educational Studies in Mathematics*, 7, 23-40.
- Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from *The American Economic Review*. *The American Economic Review*, 81, 1041-1067.
- Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht: Kluwer.
- Brown, J. R. (1997). Proofs and pictures. *British Journal for the Philosophy of Science*, 48, 161-180.
- Bundy, A. (Ed.) (2005). The nature of mathematical proof [Special Issue]. *Philosophical Transactions of the Royal Society A*, 363, 2331-2461.

- Cobb, P., & Yackel, E. (1996). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. *Educational Psychologist, 31*, 175-190.
- Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematical students. *British Educational Research Journal, 20*, 41-53.
- De Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24*, 17-24.
- De Villiers, M. (2004). The role and function of quasi-empirical methods in mathematics. *Canadian Journal of Science, Mathematics and Technology Education, 4*, 397-418.
- Dreyfus, T. (1994). Imagery and reasoning in mathematics and mathematics education. In D. F. Robitaille, D. H. Wheeler, & C. Kieran (Eds.), *Selected Lectures from the 7th International Congress on Mathematical Education* (p. 107-122). Quebec, Canada: Les Presses de l'Université Laval.
- Giaquinto, M. (2005). Mathematical activity. In P. Mancosu, K. F. Jorgensen, & S. A. Pedersen (Eds.), *Visualization, explanation and reasoning styles in mathematics* (pp. 75-87). Dordrecht, The Netherlands: Springer.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*, 93-104.
- Gowers, W. T. (2006). Does mathematics need a philosophy? In R. Hersh (Ed.), *18 unconventional essays on the nature of mathematics* (p. 182-200). New York: Springer.
- Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthèse, 134*, 45-69.
- Folina, J. (1999). Pictures, proofs, and 'mathematical practice': Reply to James Robert Brown. *British Journal for the Philosophy of Science, 50*, 425-429.
- Hanna, G. (1983). *Rigorous proof in mathematics education*. Toronto, Canada: OISE Press.

- Hanna, G. (2007). The ongoing value of proof. In P. Boero (Ed.), *Theorems in school: From history, epistemology and cognition to classroom practice* (pp. 3-18). Rotterdam: Sense.
- 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* (p. 185-212). Westport, CT: Ablex Publishing Corp.
- Harel, G., Selden, A., & Selden, J. (2006). Advanced mathematical thinking: Some PME perspectives. In A. Gutierrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (p. 147-172). Rotterdam: Sense.
- Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics III* (p. 234-282). Providence, RI: American Mathematical Society.
- Harel, G., & Sowder, L. (2005). Advanced mathematical-thinking at any age: Its nature and development. *Mathematical Thinking and Learning*, 7, 27-50.
- Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second handbook of research on mathematics education*. Greenwich, CT: Information Age Pub Inc.
- Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31, 396-428.
- Jackson, A. (2006). Conjectures no more? consensus forming on the proof of the Poincare and geometrization conjectures. *Notices of the American Mathematical Society*, 53, 897-901.
- Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: a novel class of erroneous deductions. *Cognition*, 71, 191-229.
- Jones, G. A., & Jones, J. M. (1998). *Elementary number theory*. London: Springer.

- Knuth, E. (2002). Secondary school mathematics teachers conceptions of proof. *Journal for Research in Mathematics Education*, 33, 379-405.
- Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (p. 35-60). San Diego: Academic Press.
- Littlewood, J. E. (1953). *A mathematician's miscellany*. London: Methuen.
- Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education*, 20, 41-51.
- National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.
- Peters, D. & Ceci, S. (1982). Peer-review practices of psychological journals: The fate of submitted articles, submitted again. *Behavioral and Brain Sciences*, 5, 187-255.
- Pontecorvo, C. & Girardet, H. (1993). Arguing and Reasoning in Understanding Historical Topics. *Cognition and Instruction*, 11, 365-395.
- Quality Assurance Agency. (2002). *Mathematics, statistics and operational research subject benchmark standards*. Gloucester: Quality Assurance Agency for Higher Education.
- Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, 7, 5-41.
- Recio, A., & Godino, J. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, 48, 83-99.
- Reid, D. (2002). Conjectures and refutations in Grade 5 mathematics. *Journal for Research in Mathematics Education*, 33, 5-29.
- Reid, D., & Roberts, R. (2004). Adult learner's criteria for explanations. *ZDM: The International Journal on Mathematics Education*, 36, 140-149.

- Reinard, J. C. (1988). The empirical study of the persuasive effects of evidence: The status after fifty years of research. *Human Communication Research, 15*, 3-59.
- Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (p. 89-117). San Diego: Academic Press.
- Schoenfeld, A. (1994). What do we know about mathematics curricula? *Journal of Mathematical Behavior, 13*, 55-80.
- Segal, J. (1999). Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior, 18*, 191-210.
- Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education, 34*, 4-36.
- Sternthal, B., Phillips, L. W., & Dholakia, R. (1978). The persuasive effect of source credibility: a situational analysis. *Public Opinion Quarterly, 42*, 285-314.
- Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*, 289-321.
- Stylianou, D. A. (2002). Interaction of visualization and analysis—the negotiation of a visual representation in problem solving. *Journal of Mathematical Behavior, 21*, 303-317.
- Thurston, W. P. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society, 30*, 161-177.
- Vinner, S. (1989). The avoidance of visual considerations in calculus students. *Focus on Learning Problems in Mathematics, 11*, 149-155.

- van Rooyen, S., Godlee, F., Evans, S., Smith, R., & Black, N. (1998). Effect of blinding and unmasking on the quality of peer review: A randomized trial. *Journal of the American Medical Association*, 280, 234-237.
- Walton, D. N. (1997). *Appeal to expert opinion: Arguments from authority*. University Park, PA: Pennsylvania State University Press.
- Walton, D. N. (2006). *Fundamentals of critical argumentation*. Cambridge: Cambridge University Press.
- Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, 39, 431-459.
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27, 458-477.
- Yackel, E., Rasmussen, C., & King, K. (2000). Social and sociomathematical norms in advanced undergraduate mathematics course. *Journal of Mathematical Behavior*, 19, 275-287.

Appendix

Italicized bold text was only visible to participants in the named condition.

The Heuristic Argument

Here is an open conjecture:

Conjecture. Somewhere in the decimal expansion of π there are one million sevens in a row.

Here is a heuristic argument about the claim, taken from a talk by Prof. Timothy

Gowers, University of Cambridge:

All the evidence is that there is nothing very systematic about the sequence of digits of π . Indeed, they seem to behave much as they would if you just chose a sequence of random digits between 0 to 9. This hunch sounds vague, but it can be made precise as follows: there are various tests that statisticians perform on sequences to see whether they are likely to have been generated randomly, and it looks very much as though the sequences of digits of π would pass these tests. Certainly the first few million do. One obvious test is to see whether any short sequence of digits, such as 137, occurs with about the right frequency in the long term. In the case of the string 137 one would expect it to crop up about 1/1000th of the time in the decimal expansion of π .

Experience strongly suggests that short sequences in the decimal expansion of the irrational numbers that crop up in nature, such as π , e or $\sqrt{2}$, do occur with the correct frequencies. And if that is so, then we would expect a million sevens in the decimal expansion of π about $10^{-1000000}$ of the time—and it is of course, no surprise, that we will not actually be able to

check that directly. And yet, the argument that it does eventually occur, while not a proof, is pretty convincing.

The Induction Argument

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

Here is an argument about the claim, taken from G. A. Jones & J. M. Jones,

‘Elementary Number Theory’, Springer 1998:

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

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

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

The Visual Argument

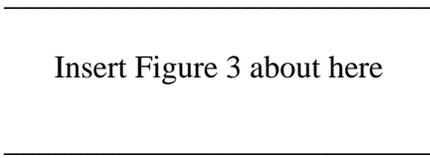
Here is a statement of the Fixed Point Theorem:

Theorem. Let $f(x)$ be continuous and increasing on $[0,1]$, such that $f([0,1]) \subseteq [0,1]$. Let $f_2(x) = f(f(x))$ and $f_n(x) = f(f_{n-1}(x))$. Then under iteration of f every point is either a fixed point, or else converges to a fixed point.

Here is an argument about the theorem, taken from ‘A Mathematician’s Miscellany’ by

Prof. J. E. Littlewood:

The only proof needed is [see Figure 3]:



The Pólya argument

Theorem. Let $f(x)$ be a real polynomial of degree $n \geq 2$ with only real roots, such that $f(x) > 0$ for $-1 < x < 1$ and $f(-1) = f(1) = 0$. Let $A = \int_{-1}^1 f(x)dx$ and let T be the area of the tangential triangle given by $f(x)$ [see Figure 4]. Then $\frac{2}{3} \cdot T \leq A$.

Here is an argument about the claim, taken from a review by George Pólya

(Mathematical Reviews, vol 1, p. 1):

Since $f(x)$ only has real roots, and none of them in the open interval $(-1,1)$, it can be written – apart from a constant positive factor which cancels out in the end – in the form

$$f(x) = (1 - x^2) \prod_i (\alpha_i - x) \prod_j (\beta_j + x) \text{ with } \alpha_i \geq 1, \beta_j \geq 1. \text{ Hence}$$

$A = \int_{-1}^1 (1-x^2) \prod_i (\alpha_i - x) \prod_j (\beta_j + x) dx$. By making the substitution $x \mapsto -x$, we find that also

$A = \int_{-1}^1 (1-x^2) \prod_i (\alpha_i + x) \prod_j (\beta_j - x) dx$. And hence by the inequality of the arithmetic and

geometric mean (note that all factors are ≥ 0):

$$\begin{aligned} A &= \int_{-1}^1 \frac{1}{2} \left[(1-x^2) \left(\prod_i (\alpha_i - x) \prod_j (\beta_j + x) \right) + (1-x^2) \left(\prod_i (\alpha_i + x) \prod_j (\beta_j - x) \right) \right] dx \\ &\geq \int_{-1}^1 (1-x^2) \left(\prod_i (\alpha_i^2 - x^2) \prod_j (\beta_j^2 - x^2) \right)^{\frac{1}{2}} dx \\ &\geq \int_{-1}^1 (1-x^2) \left(\prod_i (\alpha_i^2 - 1) \prod_j (\beta_j^2 - 1) \right)^{\frac{1}{2}} dx \\ &= \frac{4}{3} \left(\prod_i (\alpha_i^2 - 1) \prod_j (\beta_j^2 - 1) \right)^{\frac{1}{2}} \end{aligned}$$

Let us compute $f'(1)$ and $f'(-1)$. (We may assume that $f'(1), f'(-1) \neq 0$, since

otherwise $T = 0$ and the inequality $\frac{2}{3} \cdot T \leq A$ becomes trivial). By above we see that

$f'(1) = -2 \prod_i (\alpha_i - 1) \prod_j (\beta_j + 1)$, and similarly $f'(-1) = 2 \prod_i (\alpha_i + 1) \prod_j (\beta_j - 1)$. Hence we

conclude $A > \frac{2}{3} (-f'(1)f'(-1))^{\frac{1}{2}}$. Applying now the inequality of the harmonic and the geometric

mean to $-f'(1)$ and $f'(-1)$, we arrive at the conclusion

$$A > \frac{2}{3} \cdot \frac{2}{\frac{1}{-f'(1)} + \frac{1}{f'(-1)}} = \frac{4}{3} \left(\frac{f'(1)f'(-1)}{f'(1) - f'(-1)} \right) = \frac{2}{3} \cdot T.$$

Insert Figure 4 about here

Notes

¹ Some authors (e.g. Martin & Harel, 1989; Healy & Hoyles, 2000) have used the term ‘proof’ in its everyday sense, as an argument which persuades. Others (e.g. de Villiers, 1990) retain the term for deductive arguments. We have attempted to adopt the more neutral term ‘argument’ unless specifically referring to an argument which would universally be referred to by mathematicians as a proof (a more satisfactory characterization of what constitutes a proof is a difficult and current philosophical topic, and is well beyond the scope of this paper; see, for example, Azzouni, 2005; Bundy, 2005; Thurston, 1994).

² We prefer the term “gaining conviction” to more linguistically satisfying alternatives such as “becoming convinced,” as we wish to emphasize the continuous nature of one’s level of conviction. In particular, conviction in our sense does not imply a threshold level at which one moves from being unconvinced to being convinced.

³ All three universities typically required that their prospective mathematics undergraduates obtain grades of AAA—the highest possible—in their A Level examinations (the qualification taken by school leavers in England and Wales).

⁴ Both named mathematicians in Experiment 1 are very highly regarded in the mathematical community. Gowers received the 1998 Fields Medal (the mathematical equivalent of the Nobel Prize) for his work on functional analysis and combinatorics. Littlewood is well known for his prolific collaboration with Hardy in the fields of number theory and classical analysis; amongst other awards, he received the LMS De Morgan medal, the Royal Medal of the Royal Society of London and the LMS Senior Berwick Prize.

⁵ Due to an initial typographical error in the presentation of the induction argument, the first 18 responses from the researcher group were removed from the analysis of this argument.

⁶ The issue of whether or not a picture can be a proof is a controversial topic in the philosophy of mathematics literature (e.g. Brown, 1997; Folina, 1999).

⁷ Throughout the study we approached research-intensive universities from English speaking countries to participate. In practice, this meant selecting universities from the Association of Commonwealth Universities membership list (for non-U.S. universities) and the USNews.com “Best Graduate Schools” list of “top mathematics programs” (for U.S. universities).

⁸ One-tailed tests were used for the heuristic and visual comparisons, as predictions of the direction of the difference in means existed from Experiment 1 (these comparisons constituted direct replications of Experiment 1).

⁹ Note that in those comparisons where no between-conditions differences existed in Experiments 1 and 2 (the induction argument, the Pólya argument and the visual argument for students) we might expect that the two different interpretations of the instructions discussed above would lead to similar ratings. For both formal proofs we would expect high ratings on both admissibility and persuasiveness; and, as discussed above, we might expect students to give low ratings to the visual argument on both admissibility and persuasiveness.

Tables

Table 1

Mean Level of Persuasion for each Argument in Experiment 1

Group	Condition	<i>N</i>	Mean	SE
Heuristic Argument				
Researchers	Anon	94	35.8	3.13
Researchers	Named	86	45.9	3.35
Students	Anon	96	46.1	3.13
Students	Named	98	50.8	3.01
Visual Argument				
Researchers	Anon	92	36.1	3.34
Researchers	Named	98	53.4	3.70
Students	Anon	91	40.8	3.46
Students	Named	95	34.3	3.11
Induction Argument				
Researchers	Anon	89	82.2	3.24
Researchers	Named	84	80.3	3.50
Students	Anon	82	74.1	3.30
Students	Named	101	75.0	2.84

Table 2

Classification of Explanatory Comments about the Visual Argument, by Researchers and Students (collapsed across conditions) in Experiment 1

Code	Example	Res	Stu
Picture not a proof	“I was taught that a diagram is not a proof”	2	13
Closeness to proof	“The figure persuades me I could construct a more convincing proof if I wanted to”	12	0
Missing generality	“not all possibilities are included”	14	18
Other	“A lazy proof, likely to infuriate students”	7	5
Total		35	36

Table 3

Mean Level of Persuasion for each Argument in Experiment 2

Group	Condition	<i>N</i>	Mean	SE
Heuristic Argument				
Researchers	Anon	30	27.8	5.19
Researchers	Named	41	40.0	4.49
Visual Argument				
Researchers	Anon	28	39.8	6.49
Researchers	Named	39	63.5	5.03
Pólya Argument				
Researchers	Anon	28	78.7	6.16
Researchers	Named	32	78.7	4.85

Table 4

Mean Level of Persuasion for each Argument in Experiment 3

Group	Condition	<i>N</i>	Mean	SE
Heuristic Argument				
Researchers	Anon	95	24.2	2.83
Researchers	Named	119	35.8	3.13
Visual Argument				
Researchers	Anon	106	36.2	3.53
Researchers	Named	104	47.6	3.29
Induction Argument				
Researchers	Anon	91	90.4	2.19
Researchers	Named	115	83.9	2.81

Figures

Figure 1. Mean level of persuasion for each argument in Experiment 1. Error bars represent ± 1 SE of the mean.

Figure 2. An illustration of the proposed model.

Figure 3. The image used in the induction argument.

Figure 4. The image used in the Pólya argument.

Figure 1. Mean level of persuasion for each argument in Experiment 1. Error bars represent ± 1 SE of the mean.

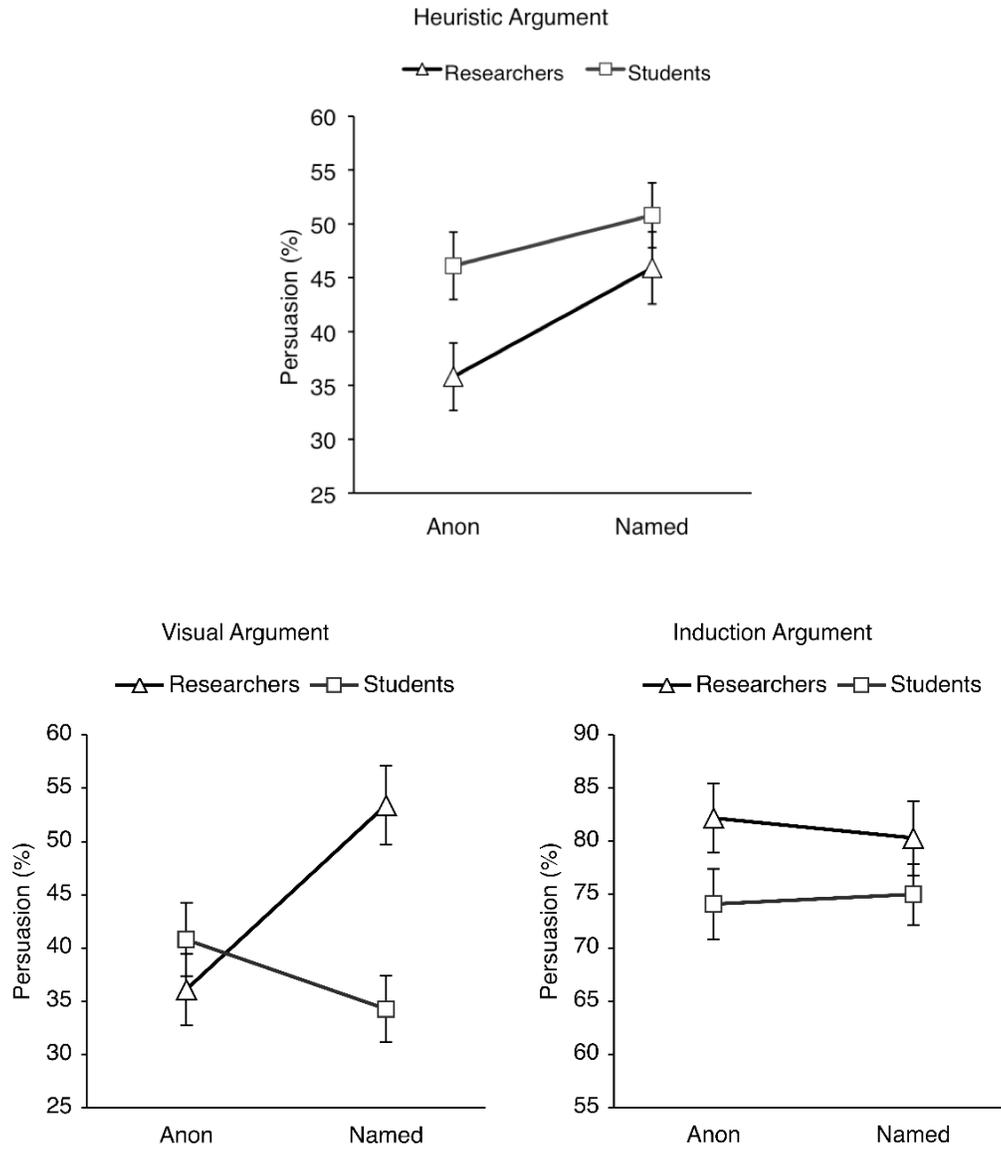


Figure 2. An illustration of the proposed model.

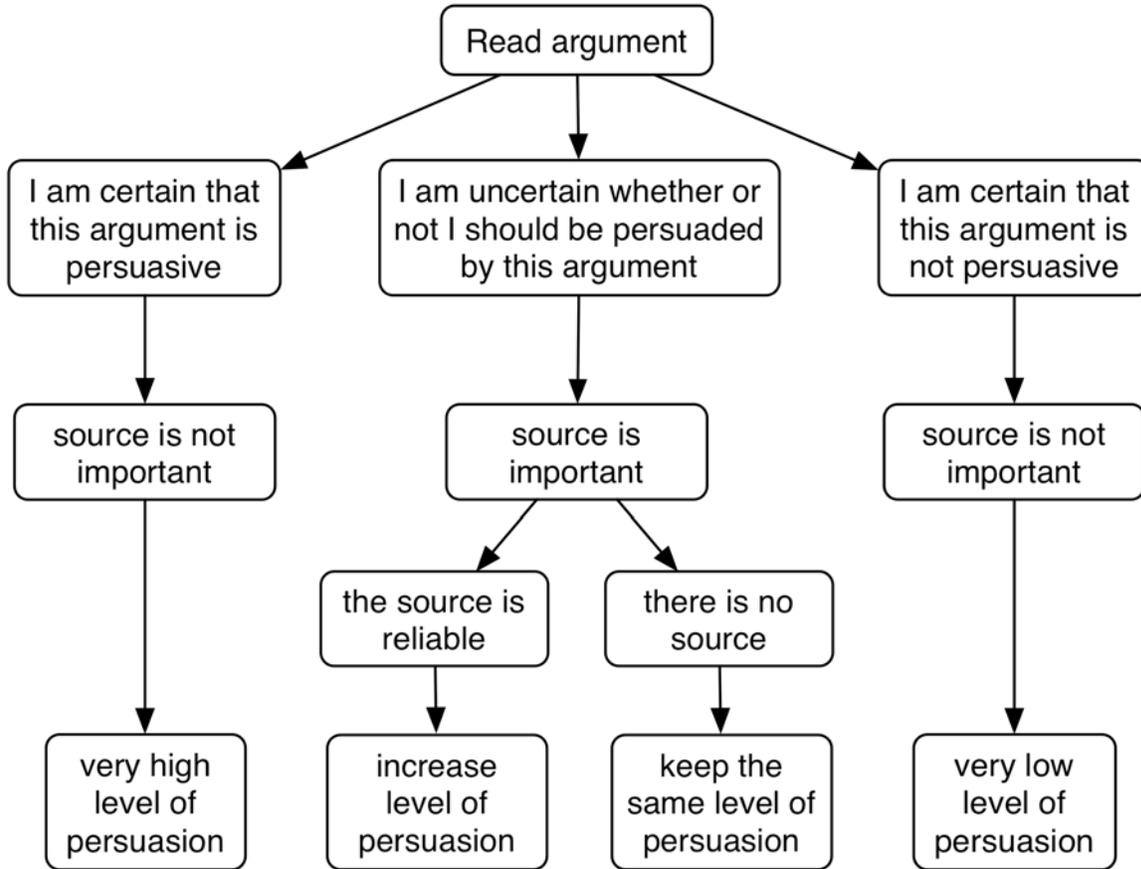


Figure 3. The image used in the induction argument.

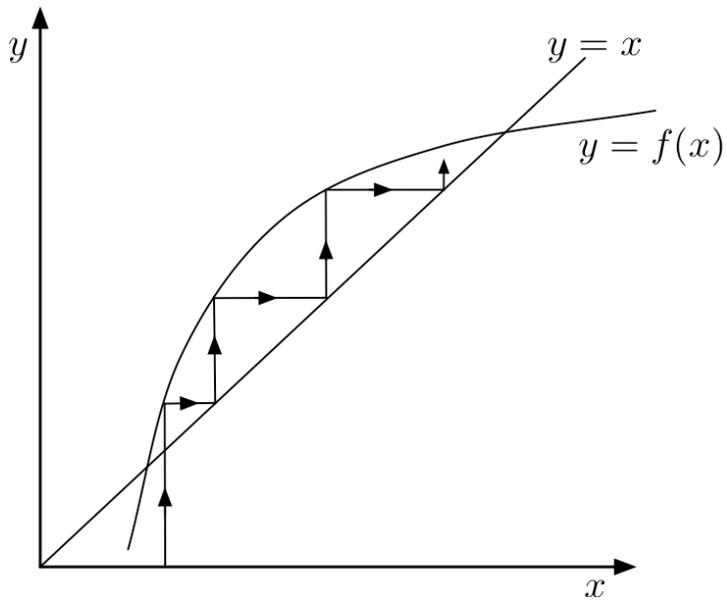


Figure 4. The image used in the Pólya argument.

