IS IT EVER APPROPRIATE TO JUDGE AN ARGUMENT BY ITS AUTHOR?

Matthew Inglis and Juan Pablo Mejia-Ramos

University of Warwick

There is a widespread belief in the mathematics education community that students should be encouraged to avoid basing their level of conviction in mathematical arguments on the authority of the argument's source. In this paper we report an experiment which investigated the role that authority plays in the argument evaluation strategies of undergraduate students and research active mathematicians. Our data show that both groups were more persuaded by an argument if it came from an authority figure. The implications of this finding are discussed. It is argued that the role of authority in mathematical argumentation – both in terms of actual behaviour and of normative behaviour – requires deeper scrutiny.

AUTHORITY IN ARGUMENTATION

In day-to-day reasoning it is well known that the source of an argument has a significant impact on how persuaded an audience will be. For example, when judging the persuasiveness of an argument about Avian Bird Flu, an audience is likely to react differently if it comes from a government scientist compared to a taxi driver. In mathematics, however, it is often claimed that the source of a proof should be ignored. Selden and Selden (2003), for example, wrote:

> "Like midcentury structural critics, mathematicians seem to treat a proof as being independent from its author." (Selden & Selden, 2003, p.6).

As part of their 'proof scheme' framework, Harel and Sowder (1998) defined a person's proof scheme as that which "constitutes ascertaining and persuading for that person" (Harel & Sowder, 2005, p.33). In a large study looking at what types of proof schemes students have, they found widespread evidence of the so-called *authoritarian proof scheme*. A person with such a scheme becomes persuaded of a statement's truth or falsity based on the pronouncements of an authority figure, often a teacher or textbook. Harel and Sowder (2005, p.42) described the authoritarian proof scheme as "an undesirable, yet common, way of thinking" and argued that "[instruction] must institute a didactical contract that attempts to suppress the authoritarian proof scheme" (Harel & Sowder, in press).

These negative comments about the authoritarian proof scheme seem to be based upon *a priori* views of expert practice which have not been verified by empirical research. In this paper we report a brief experimental study which empirically investigates how the presence of an authority figure influences conviction in mathematical argumentation. However, it is important to emphasise that, in this paper, we are not investigating proof in the sense of Selden and Selden (2003). Instead we advocate a broader framework for studying argumentation in mathematics that includes both formal proof and less rigorous arguments (e.g. Inglis & Mejia-

Ramos, 2006; Inglis, Mejia-Ramos & Simpson, submitted). Crucially, the experimental instrument discussed in this paper does *not* involve an argument that carries *absolute* conviction, and so direct comparisons with Harel and Sowder's work are potentially misleading. Notwithstanding this caveat, we believe that studying the factors which influence experts' conviction in mathematical arguments does have implications for the received view of the authoritarian proof scheme.

A CONJECTURE, AN ARGUMENT, AND AN EXPERT

Materials

Our task consisted of the following conjecture, argument and question:

> *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:*
>
> 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.
>
> *After having read this argument please say to what extent you are persuaded by it:*
>
> not persuaded 1 2 3 4 5 totally persuaded

The argument and conjecture was taken from a book chapter written by Professor Timothy Gowers, a highly respected Fields Medal winning mathematician from the University of Cambridge (Gowers, 2006).

Method

We asked participants to complete the task above. Each participant was randomly assigned into one of two groups, the anonymous group and the named group. The anonymous group's task contained the line

> *Here is a heuristic argument about the claim.*

Whereas the named group's task contained the line:

> *Here is a heuristic argument about the claim (taken from a talk by Prof. Timothy Gowers, University of Cambridge).*

Participants completed the task online. The undergraduate students (*N*=302) were all from a single highly-rated university and were asked to participate by means of an email from the departmental secretary. The email explained the task and asked them to click through to the experimental website should they wish to participate. The research active mathematicians were recruited in two different ways. Some (*N*=14) were staff members at the same university as the undergraduates and were recruited in a similar manner. Others (*N*=60) were recruited through an advertisement posted on a mathematics research newsgroup. Before taking the task, participants in the researcher group were asked to self-declare that they were research active mathematicians.

The number of psychological studies conducted using the internet has dramatically increased in recent years, and the method obviously raises important questions regarding reliability and, in particular, possible multiple-submission. To prevent this, participants' IP addresses were recorded and analysed to detect possible rapid resubmissions. Space constraints prevent a full discussion of the reliability and validity of internet research here, although see Gosling, Vazire, Srivastava, & John, (2004) or Reips (2000).

Participants' comments on the argument

After participants had rated their level of persuasion in the argument, they were given the opportunity to write comments to explain their selection. The comments from the research active mathematicians emphasised the lack of agreement about how to react to the argument. The following selection of comments are typical of the range:

> Purely logically on the basis of the evidence presented, I am not persuaded at all. (Researcher, Anonymous Group, Rated 2).

> Normalcy of (the digits) of π is not unreasonable given almost all reals are normal. Then again, almost all real numbers don't have names, so who knows… (Researcher, Anonymous Group, Rated 2).

> It is persuasive but not a proof, but it only claims to be heuristic. (Researcher, Anonymous Group, Rated 5).

> The argument does not in any way claim to be a proof of the conjecture. […] The argument is actually for a much stronger result than the conjecture, and the conjecture could be true even if the argument is false. On balance, then, I would bet money that the conjecture is true. (Researcher, Named Group, Rated 4).

Along with there being no agreement as to whether the argument was persuasive or non-persuasive, a clear majority of both the undergraduates and the researchers rated their level of persuasion as being neither totally persuaded nor totally unpersuaded (in the range 2-4). This observation supports Inglis et al.'s (submitted) assertion that a key part of developing as a mathematician is the construction and appropriate qualification of arguments with non-deductive warrants.

Results

The mean level of persuasion from each of the groups are shown in Figure 1.

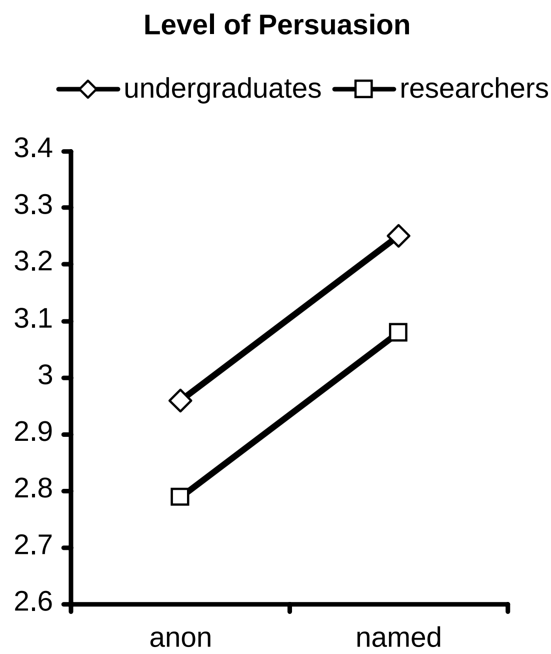

Figure 1: The mean levels of persuasion for each group for each version.

We conducted two nonparametric between-groups comparisons to identify possible main effects. There was no significant effect for group-type (undergraduate vs. researcher), Mann Whitney U=14052, *NS*. There was, however, a significant main effect for task-type (anon vs. named), Mann Whitney U=17498, p=0.014, with the group who knew the author's identity ranking the argument as more persuasive than the anonymous group.

Figure 1 clearly indicates that there was no task×group interaction, but to formally test for this we conducted an ordinal logistic regression analysis, resulting in a ordered logit model. The model was a poor fit, R^2=0.02, but the overall relationship was significant, $\chi^2(6)$=18.4, p=0.031. As the Mann Whitney tests suggested, in the model there was a significant effect for task-type, $\chi^2(1)$=3.95 p=0.041, and not for group-type, $\chi^2(1)$=0.347, *NS*. As expected, no evidence was found of a significant task×group interaction, $\chi^2(1)$=0.011, *NS*.

The results of these analyses are clear: participants in the study were more likely rate the argument as being persuasive if they were aware that it came from an authority figure. Furthermore, this effect was found in both undergraduates and research active mathematicians.

DISCUSSION

The data from this study indicate that when evaluating their level of conviction in non-deductive arguments, both undergraduates and research active mathematicians are influenced by the presence of an authority figure. Given the comments of Harel and Sowder (1998, 2005, in press) reported above, this result is perhaps surprising. However, these authors were referring to arguments which carry absolute conviction, not non-absolute arguments of the type used here.

There are also several potential differences between the argument used in this paper and a traditional proof. For one, some parts of Gower's argument need to be taken on trust. He asserts that statisticians have performed various tests for normality, but does not quote references or test outcomes. Perhaps participants who were not aware of the author's identity would be more likely to disregard this factual information as untrustworthy than those who knew Gower's status. Another possibility is that participants who were aware of the author's reputation would be reassured that he was not trying to hide evidence which would weigh against his argument, indeed one participant explicitly stated how knowledge of the author reassured him/her of this:

> 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, Named Group, Rated 4).

However other participants made comments which suggested that the way they react to an argument (including proofs) does indeed vary depending on the identity of the author. One researcher commented that if the claim being made is sensible ("ordinary-looking") they would be less likely to devote time to checking the claim's justification if it came from a reputable source than they would if the author was unknown or unfamiliar. Ecologically this makes a lot of sense, disregarding the source of an argument when evaluating one's level of conviction in it involves giving up a lot of potentially useful information. Arguments proposed by reputable sources are more likely to be sound than those proposed by people with a history of putting forward erroneous claims. Throwing away helpful data when judging such matters would, in most domains, be regarded as irrational.

Notwithstanding the differences we acknowledge between the task used in this paper and formal mathematical proof, and between arguments that provide an absolute level of conviction and those that provide intermediate levels of conviction, we suggest that the data reported here indicates that the role of authority in mathematical argumentation (including proof) requires deeper scrutiny. Within a framework that models degrees of conviction in argumentation appropriate normative standards for the role that authority should play are unclear. In some cases it seems entirely reasonable to base conviction in a mathematical claim on an authority figure, when applying a published theorem to a new situation, for example; in other situations it seems less reasonable: most educators would not want to encourage their students to base their levels of conviction simply on the word of the teacher.

We suggest that existing descriptions of the role of authority in mathematical argumentation (and proof) are too imprecise. Harel and Sowder (1998, 2005, in press) suggest that basing conviction on authority is “undesirable” as it is not how professional mathematicians behave. This paper demonstrates that this need not be the case for non-deductive arguments. We suggest that the issue of whether it is the case for deductive proof remains an open question, one suitable for further empirical research.

REFERENCES

Gosling, S. D., Vazire, S., Srivastava, S., and 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.

Harel, G., and Sowder, L. (1998). Students’ proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, and E. Dubinsky (Eds.), *Research in Collegiate Mathematics III* (p. 234-282). Providence, RI: American Mathematical Society.

Harel, G., and Sowder, L. (2005). Advanced mathematical-thinking at any age: Its nature and development. *Mathematical Thinking and Learning*, *7*, 27-50.

Harel, G., and Sowder, L. (in press). Towards comprehensive perspectives on the learning and teaching of proof. To appear in F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning.* NCTM.

Inglis, M., and Mejia-Ramos, J. P. (2006, June). Applying informal logic to arguments in mathematics. Presented at the International Conference for the Teaching of Mathematics at the Undergraduate Level, Istanbul, Turkey.

Inglis, M., Mejia-Ramos, J. P., and Simpson, A. (submitted). Modelling mathematical argumentation: The importance of qualification. *Submitted for publication, draft available upon request.*

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.

Selden, A., and 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*(1), 4-36.