On the persuasiveness of visual arguments in mathematics[†]

Matthew Inglis
Mathematics Education Centre, Loughborough University.

Juan Pablo Mejía-Ramos
Institute of Education, University of Warwick.

Abstract. Two experiments are reported which investigate the factors that influence how persuaded mathematicians are by visual arguments. We demonstrate that if a visual argument is accompanied by a passage of text which describes the image, both research-active mathematicians and successful undergraduate mathematics students perceive it to be significantly more persuasive than if no text is given. We suggest that mathematicians' epistemological concerns about supporting a claim using visual images are less prominent when the image is described in words. Finally we suggest that empirical studies can make a useful contribution to our understanding of mathematical practice.

Keywords: belief, conviction, experimental methods, persuasion, visual arguments

1. Introduction.

In recent years there has been a growing interest among philosophers of mathematics about the practice of mathematicians, and especially practice related to argumentation. Corfield (2003), for example, complained that "By far the larger part of the activity which goes by the name of philosophy of mathematics is dead to what mathematicians think or have thought" (p. 5). Instead, he argued that philosophers should pay much closer attention to the actual practice of mathematicians. Although this suggestion to focus upon mathematical practice has been taken up by subsections of the philosophical community (e.g. Mancosu et al., 2005; Van Kerkhove and van Bendegem, 2006), to date there have been relatively few empirical studies of mathematical practice.

This lack of empirical studies is somewhat surprising as in recent years there has been growing interest in applying empirical research methods to philosophical questions. In particular experimental methods have been widely used to systematically explore 'folk intuitions' of ethical dilemmas (e.g. Appiah, 2008; Nadelhoffer and Nahmias, 2007). Understanding mathematical practice would seem to be an area even more suited to the application of empirical methods: after all, mathematical practice essentially refers to the behaviour of mathematicians in mathematical situations; and behaviour is essentially an empirical matter.

Our primary goal in this paper is to argue that empirical studies of mathematical behaviour can give insights into mathematical practice which are not

[†] This is a non-final preprint, to appear in *Foundations of Science*.

easily glimpsed using other methods. In short, we will attempt to demonstrate that empirical studies can make a useful contribution, alongside more traditional philosophical analyses, to our understanding of mathematical practice. To achieve this aim we first review earlier empirical work which has studied the argumentation behaviour of undergraduate mathematics students. We then report the results of a study which interrogated one particular factor that influences how persuaded research-active mathematicians and undergraduate mathematics students are by visual mathematical arguments: the presence or absence of descriptive text.

2. Background.

2.1. ARGUMENTATION IN MATHEMATICS EDUCATION.

Our goal in this section is to briefly review earlier empirical work on mathematical argumentation and proof. The focus is on the argumentative activities of undergraduate students, primarily because, to date, very few studies have looked at the activity of research-active mathematicians. Empirical work in this area has focussed on two different activities: the construction of novel arguments, and the reading of given arguments. We consider each of these activities in turn.

Harel and Sowder (1998) introduced the framework of *proof schemes*, which they defined as the different ways in which undergraduate students gain certainty about, and persuade others of, the truth of mathematical statements. Harel and Sowder found that students primarily construct three types of arguments: those relying on features that are external to the students (e.g. the form of the argument, an authority figure, or meaningless symbolic manipulation), empirical arguments using different types of examples (e.g. visuo-spatial images, numerical substitutions, measurements), and deductive arguments that range from those expressed in terms of generic examples to those in which students exhibit some understanding of their argument's dependence on an given axiomatic system (Harel, 2007; Harel and Sowder, 2007).

Although Harel and Sowder's (1998) association of the activities of (i) gaining personal certainty about a statement's truth, and (ii) persuading others of a statement's truth may be reasonable when analysing groups of people who are persuaded by similar argumentation methods, it is less clear when this is not the case. For example, a student might use her own methods to persuade other students, but might deploy other methods to persuade a teacher who is known to value other types of argumentation. Evidence supporting this possible disassociation was provided by Healy and Hoyles (2000), who conducted a large survey of high-attaining secondary school students' notion of proof in algebra and found a significant difference between the type of arguments students would use to support a given claim (mostly empirical arguments) and the type of arguments which students thought would receive the best mark from their teachers (mostly formal algebraic arguments). One interpretation of these findings in terms of the two subprocesses constituting a proof scheme is that the students in Healy and Hoyles's study used inductive methods in the process of convincing themselves, but symbolic methods in the process of persuading their teachers, i.e. the two components of a proof scheme were disassociated in that particular context.

Harel and Sowder's (1998) sole focus on arguments which give certainty to a conclusion has also been questioned. During an investigation of the ways in which research students evaluate conditional statements, we found that students construct the different types of arguments in Harel and Sowder's (1998) taxonomy, not solely with the intention of gaining *certainty*, but often to merely *reduce* their doubts about the truth of a statement (Inglis et al., 2007). The talented mathematics research students were willing to construct (often highly sophisticated) empirical and intuitive arguments to *reduce* their doubts regarding a particular claim, consistently pairing these types of justification with non-absolute qualifiers like "it is probable that", "it is plausible that" and "it seems that".

Recio and Godino (2001) interpreted Harel and Sowder's (1998) notion of proof scheme as referring to the type of arguments that students would construct when asked to 'prove' a given statement (although Harel and Sowder's data were actually drawn from a wider range of task-types than this). They found that while some students constructed deductive arguments (some informal and some formal), and some constructed empirical arguments (some implying generality, some not), they tended to construct the same type of argument across different tasks, suggesting that proof schemes may be relatively stable across different mathematical content areas.

Other researchers have focussed on the methods by which students construct deductive arguments, finding that some construct arguments based on heuristic ideas (involving examples and visual representations) and others base their arguments on procedural ideas (involving formal symbolic manipulations) (Raman, 2003; Weber and Alcock, 2004). Again, there is evidence that some students consistently adopt one of these two approaches (Duffin and Simpson, 1993; Pinto and Tall, 2002).

Another strand of research has focused on students' reading of mathematical arguments. In particular, on the ways in which they determine whether or not a given argument is a mathematical proof (including the types of arguments they are willing to call a proof), and on the ways in which they evaluate different types of arguments against a wider variety of criteria (e.g. personal conviction, general persuasiveness, display of understanding, validity, marks it would get in an exam).

Martin and Harel (1989) established that nearly half of the students they surveyed were willing to give high "proof ratings" to both deductive and inductive arguments justifying the same claim. They interpreted these findings as evidence that students needed both types of arguments to gain conviction. Other studies have found that, when asked to evaluate arguments with respect to personal conviction and validity, many students tend to consider empirical arguments as personally convincing but invalid, and purported proofs as personally convincing and valid, regardless of their actual validity (Raman, 2002; Segal, 2000).

Selden and Selden (2003) found that without intervention from an interviewer, students did not perform better than chance at recognising a correct proof from other invalid arguments with proof-like characteristics. In contrast to the appropriate ways in which students *say* that they read proofs, Selden and Selden found that students' criteria for deciding whether or not an argument is a proof involved local aspects of the argument (which led them to detect minor/superficial mistakes instead of more global ones) and whether or not they understood the argument (leading them to accept as proofs arguments they could easily understand, regardless of whether or not these arguments actually proved the given claim). However, they also noted that when these students were encouraged to reflect on their decisions, the percentage of students' correct judgements rose considerably (cf. Alcock and Weber, 2005). In a rare study involving mathematicians, Weber (2008) observed that researchers would sometimes want to know more about the context from which the proof was taken (i.e. whether it was intended for first year or advanced undergraduates) before judging its validity. He also noted that his participants would sometimes use non-deductive methods when proof validating, that is to say that they would use empirical or informal methods to determine whether one statement in a proof followed from the previous one.

Weber's (2008) study was unusual in that it focussed upon the argumentation activities of research-active mathematicians. Although, to date, there has been little empirical work in this area, we believe that such studies have the potential to make important contributions to philosophical discussions of mathematical practice. To illustrate this point, in the remainder of the paper we describe a study which interrogated the factors that influence how persuasive mathematicians and undergraduate students find visual mathematical arguments. We first briefly review earlier discussions on the epistemological status of such arguments.

2.2. Philosophy of visual arguments.

In recent years some philosophers of mathematics have become increasingly interested in the status of visual arguments (arguments that rely upon pictures – including not only drawn or computer-generated figures and diagrams, but also mental images) in mathematical practice. In particular, discussions have focussed on the relationship between visual arguments on the one hand, and mathematical proof and discovery on the other (Barwise and Etchemendy, 1991; Brown, 1999; Dove, 2002; Mancosu et al., 2005; Giaquinto, 2007). In general, these discussions at some point present what is considered to be the *common view* of the role that visual representations play in mathematics: i.e. that pictures may be useful heuristic tools which suggest ways of understanding proofs, but that they are nevertheless inappropriate when it comes to providing unequivocal, reliable evidence to support a mathematical claim, let alone providing a proof. The origin of this view is often traced back to late nineteenth century mathematicians, who having encountered false mathematical claims that seemed to be obviously true on account of visual arguments, learned to distrust pictures and avoid relying on visualisation. For some authors visual arguments are not always inappropriate, and they argue that, at least in some cases, visual thinking can be reliable and hence deliver knowledge (e.g. Giaquinto, 2007). Others go as far as to claim that in some cases pictures on their own can be fully-fledged proofs (e.g. Brown, 1999).

In this paper, we do not address the difficult normative question of whether a picture should be used as an integral part of a proof (i.e. whether arguments that *rely* on pictures could be taken as proofs), or whether pictures should retain their status as heuristic adjuncts to proofs (i.e. whether proofs must be, at least in principle, completely independent of any picture). A brief review of these different philosophical positions on the status of pictures in proofs was presented by Hanna and Sidoli (2007). Rather, we focus on a particular aspect of this discussion: the formation of belief from visual arguments.

Few empirical studies have been conducted about the persuasiveness of visual mathematical arguments. However, our recent study of the role of authority in mathematics did touch on the subject (Inglis and Mejía-Ramos, 2007). We gave two groups of participants – research-active mathematicians and successful mathematics undergraduates – a visual 'proof' of the one-dimensional fixed point theorem (taken from Littlewood, 1953, p. 37), and asked them to rate how persuaded they were by the argument. Half of the participants were told that the argument had been written by Littlewood, whereas the remaining half were not.

The research-active mathematicians in the sample were strongly influenced by the authority figure: those who saw Littlewood's name rated the argument as being significantly more persuasive than those who did not (a difference of close to 20 percentage points). Surprisingly, the undergraduate students in the study showed a slight non-significant trend in the opposite direction. To explore this difference in more detail we conducted an analysis of the explanatory comments left by participants, finding that undergraduate students tended to be more dogmatic about the epistemological acceptability of visual arguments than their research-active colleagues. A typical undergraduate comment was of the form "I was taught that a picture is not a proof", whereas researchers tended to concentrate more on how easy it would be to turn the picture into a formal proof. One researcher, for example, wrote "I could construct a more convincing proof if I wanted to".

It seems then, that for at least some groups of participants, the author of a visual argument can impact upon its persuasiveness (the extent to which it provides support to its conclusion). In this paper we focus upon one further factor which may influence the persuasiveness of visual mathematical arguments: the presence or absence of descriptive text. Hanna and Sidoli (2007) pointed out that in the philosophical debate regarding the role of visual arguments in mathematics some authors give particular importance to the verbal/symbolic text that accompanies the picture, which may be seen as the result of "extracting" the information implicitly presented in the visual representation. Here we present two experiments which were designed to interrogate the effect of the presence or absence of a passage of descriptive text on mathematics undergraduates' and research-active mathematicians' evaluations of a visual argument.

3. Experiment 1.

3.1. Method.

The experiment we report here was conducted online, using code produced by the WEXTOR system (Reips and Neuhaus, 2002). The validity and reliability of internet methods have been extensively discussed in the research methods literature (e.g. Gosling et al., 2004; Reips, 2000). A particularly serious threat to validity comes from the possibility of participants submitting multiple responses. To deal with this issue we took two steps. First, in line with established practice (e.g. Klauer et al., 2007) we kept a record of participants' internet protocol (IP) addresses, and only recorded the first submission from each address. Second, we issued an explicit instruction on the first page of the experiment warning participants that engaging in such practices would be detrimental to the research.

3.1.1. *Participants and procedure.*

Participants came from two groups: undergraduate students and research-active mathematicians. The students ($N = 58$) were from three highly ranked UK universities. The research-active mathematicians ($N = 56$) worked at universities in the United States. Both groups participated without payment and were recruited via an email from their departmental secretary. The email explained the purpose of the experiment, and asked recipients to visit the experimental website if they wished to participate. Once at the website, they were asked to declare that they were a "research active mathematician" or an "undergraduate student" before commencing the task. Data from each group was collected from separate addresses several weeks apart to ensure the integrity of the data. A few participants who appeared to make false declarations (i.e. they declared themselves to be researchers during the undergraduate data collection period) were removed from the analysis. After the declaration participants were randomly assigned into one of two experimental conditions: 'description' and 'no-description'.

A screen of instructions (given in full in the Appendix) was displayed followed by the first of six trials. Each trial consisted of a *claim*, some *evidence* about the claim, and the following instructions:

> Your task is to evaluate the extent to which the given evidence, and only the given evidence, provides support to the following claim.

Participants were required to respond to the trial using the form shown in Figure 1. The aim of this question was to prompt participants into conducting what we have called elsewhere a Type 2 evaluation (Inglis and Mejía-Ramos, 2008).[1] In terms of Toulmin's (1958) argumentation scheme, shown in Figure 2, a Type 2 evaluation revolves around determining what type of modal qualifier you would be prepared to pair with the given data, warrant, backing and conclusion. In the present context the aim was to encourage participants to construct an argument whose conclusion was the given claim, whose data was the given evidence, to infer an appropriate warrant and backing, and then to determine which of the given options was the most appropriate modal qualifier (see Figure 2).

In addition to the tick-box responses, introspective comments were invited via a text response box. Once participants had selected their response, they clicked submit, and the next claim/evidence pair was loaded. The order in which the pairs appeared was randomised for each participant.

3.1.2. *Materials.*

Participants saw a total of six claim/evidence pairs. The evidence offered was of different types: two trials had empirical evidence, two had visual evidence (one of which was the experimental trial) and two had evidence based on the authority of the publication place. The non-experimental trials were unrelated

[1] In earlier work we suggested that, when asked to rate how persuaded they are by a given argument, a participant might focus their evaluation on one of five areas (Inglis and Mejía-Ramos, 2008). They could focus on: (0) the *data* of the given argument and how significant/trustworthy it is; (1) the likelihood of its *conclusion*; (2) the strength of the *warrant* (and its associated *backing*); (3) the given *qualifier* (and its associated *rebuttal*) and the extent to which this qualifier is appropriate considering the rest of the argument; and (4) a particular *context* in which the given argument may take place. In turn, each of these foci of attention can (and, in the examples of participants' evaluations we discussed, did) provide a different type of response (which we labelled Types 0-4, correspondingly).

Please tick one of the options. The given evidence suggests that the claim is:

◯ definitely true
◯ almost certain to be true
◯ very likely to be true
◯ more likely to be true than not; OR
◯ this gives no useful information about the claim.

Figure 1. The form by which participants responded to the task.

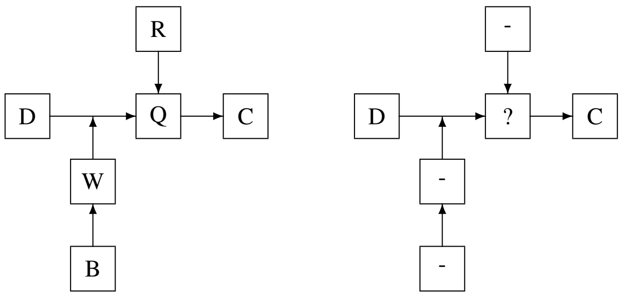

Figure 2. Left: Toulmin's model of a general argument, showing the data (D), warrant (W), backing (B), modal qualifier (Q), rebuttal (R) and conclusion (C) components. Right: The layout of the argument in the current experiment. The data and conclusion were given, and participants inferred the warrant, backing and then made a judgement about the appropriate modal qualifier.

to the issues discussed in this paper and can be seen as distractor or filler tasks, designed to disguise the purpose of the experiment, and to provide a 'high hurdle' to participation.

The claim used for the experimental trial was a statement of Young's Inequality:

> *Claim:* Let ϕ and ψ be two continuous, strictly increasing functions. Suppose $\phi = \psi^{-1}$ and $\phi(0) = \psi(0) = 0$. Then, for $a, b \geq 0$, we have:
>
> $$ab \leq \int_0^a \phi(x)\,dx + \int_0^b \psi(y)\,dy$$
>
> with equality if and only if $b = \phi(a)$.

For the no-description condition, the evidence consisted of two images, shown in Figure 3. In the description condition, the evidence consisted of the same two images followed by some descriptive text:

> With the given conditions, the graph of $\phi(x)$ coincides with the 'rotated' graph of $\psi(y)$ (plotted with the independent variable in the y-axis, and the dependent variable in the x-axis). Therefore, the value of $\int_0^a \phi(x)\,dx$ equals the area, from 0 to a, between this graph and the x-axis, while the value of $\int_0^b \psi(y)\,dy$ equals the area, from 0 to b, between this graph and the y-axis. Notice that for $a, b \geq 0$, the sum of these areas is greater than, or equal to ab (the area of the rectangle of sides a and b).

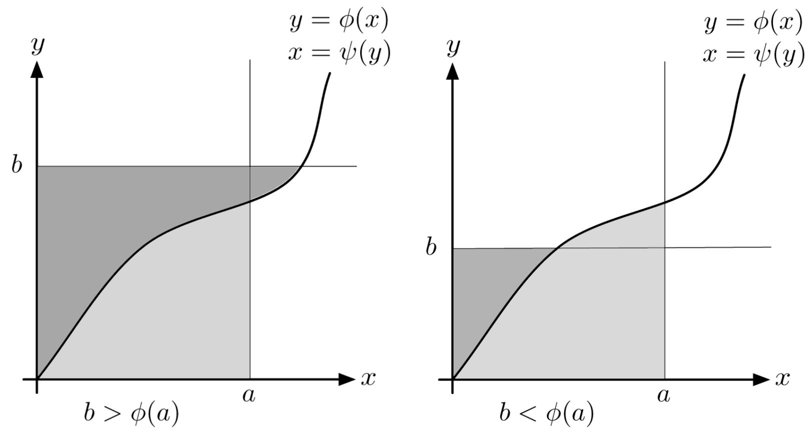

Figure 3. The images used as evidence in Experiment 1.

Our aim in choosing this text was to describe the image in words but, crucially, not to provide any deductive argument in support of the claim.

3.2. RESULTS.

The data analysis was conducted using the Scientific LogAnalyzer method (Reips and Stieger, 2004). There were large individual differences in participants' responses. One mathematician in the description condition, for example, rated the argument as 'definitely true' and wrote

> "The proof seems convincing, ... and very nice!"

Another, in the no-description condition, rated the argument as 'very likely to be true' and wrote

> "The general case has not been proved."

The overall breakdown of responses are shown in Figure 4 and Table I. Over half of the participants in the description condition rated the evidence as indicating that the claim was definitely true; whereas less than a quarter of the participants in the no-description condition chose this option. Ten participants in the no-description condition claimed that the evidence offered no useful information about the claim, compared to only four in the description condition. Participants' responses in the two conditions were analysed using a Mann Whitney test. It was found that those in the description condition judged that the argument provided significantly more support to the claim than those in the no-description condition, $U = 964, p < 0.001$. When the research-active mathematicians and undergraduate students were analysed separately, this effect retained significance; mathematicians, $U = 208, p = 0.001$; students, $U = 240, p = 0.006$.

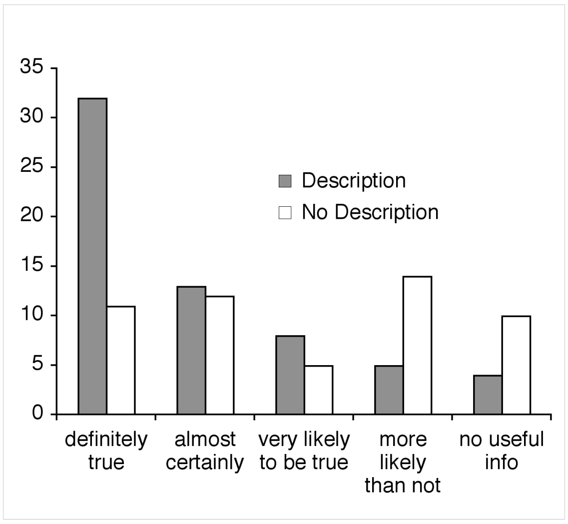

Figure 4. The number of participants selecting each response, from both groups, in Experiment 1.

Table I. Responses from participants in the description (desc) and no-description (no-desc) conditions in Experiment 1.

	Researchers		Students	
	desc	no-desc	desc	no-desc
definitely true	19	9	13	2
almost certainly	7	7	6	5
very likely to be true	2	3	6	2
more likely than not	0	5	5	9
no useful info	0	4	4	6

In discussions regarding the degree of belief conferred on mathematical claims on account of visual evidence, it seems to be particularly important to determine whether visual arguments simply suggest that the claim may be true (in which case this kind of evidence could only be used to establish conjectures), or whether visual arguments can in fact remove all doubts and provide certainty about the truth of the claim (cf. the proof schemes framework; Harel, 2007; Harel and Sowder, 1998).

In view of the importance of the distinction between arguments which remove doubts and those which merely reduce doubts, we reclassified participants' responses. For those participants who chose the 'definitely true' option, the given evidence had provided them certainty that the claim was true, while for those who chose one of the remaining options, the evidence had left them with some level of uncertainty. Recoding participants' responses into two categories – 'definitive' (definitely true) and 'heuristic' (all other responses) – revealed that those in the description condition were more likely to consider the evidence definitive than those in the no-description condition, $\chi^2(1) = 11.2, p = 0.001$. Again, this effect retained significance when the mathematicians and students were analysed separately; mathematicians, $\chi^2(1) = 7.1, p = 0.008$; students, $\chi^2(1) = 6.6, p = 0.010$.

3.3. DISCUSSION.

Both the research-active mathematicians and the undergraduate students who participated in Experiment 1 found the visual argument about Young's inequality to be more persuasive when it was accompanied by a passage of descriptive text. There would seem to be two reasonable hypotheses to account for these data. One account is that mathematicians tend to be unwilling to use evidence which consists solely of visual images to gain high levels of conviction in mathematical statements; but that when that evidence is accompanied by descriptive text these concerns are less prominent as the argument is less obviously entirely visual. However, an alternative account might be that some of the participants in the no-description condition were unsure of what to attend to in the visual image: that the lack of a descriptive text made the image hard to relate to the claim. To rule out this possibility we conducted Experiment 2.

4. Experiment 2.

Experiment 1 demonstrated that visual arguments can be perceived as more persuasive when accompanied by descriptive text than they are when left without a description. The main goals of Experiment 2 were to (i) replicate this finding; and (ii) to rule out the suggestion that the descriptive text merely helps participants attend to salient features of the argument.

4.1. METHOD.

Participants were research-active mathematicians ($N = 24$) employed in Australian universities, and mathematics undergraduates ($N = 39$) from two highly ranked UK universities. As with Experiment 1, participants were contacted by email via their departmental secretary.

The method was kept identical to that used in Experiment 1, with one exception. Again, six tasks were used, five of which were direct replicas of the tasks used in Experiment 1. The remaining task – the Young's Inequality trial – had an identical description condition to Experiment 1, but the

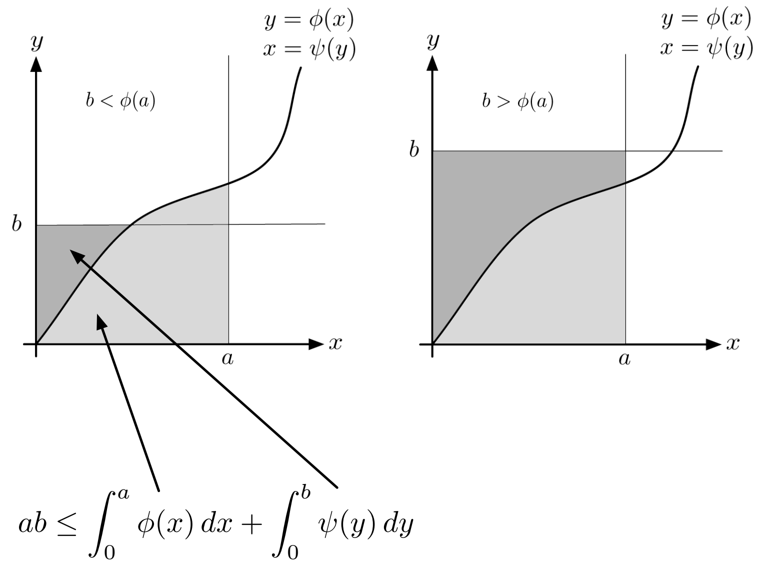

Figure 5. The image used as evidence in the no-description condition of Experiment 2.

no-description condition (here called the arrows condition) used a different image, shown in Figure 5. This modified image was designed to direct attention (using arrows) towards salient features of the visual argument. If the primary reason for the effect detected in Experiment 1 was that the description directed participants' attention towards salient parts of the visual argument, then we would expect that the effect would be abolished in Experiment 2.

4.2. Results and discussion.

The breakdown of responses are shown in Figure 6. Participants' responses in the two conditions were analysed using a Mann Whitney test; it was found that those in the description condition judged that the argument provided significantly more support to the claim than those in the arrows condition, $U = 288, p = 0.016$. As before, responses were recoded into 'definitive' (definitely true) and 'heuristic' (all other responses). Participants in the description condition were more likely to categorise the evidence as definitive compared to participants in the arrows condition, $\chi^2 = 5.26, p = 0.022$.

The results of Experiment 2 essentially replicated the effect found in Experiment 1: participants who saw the descriptive text reported that they found the evidence to be more convincing than those who saw the image with attention-directing arrows. Consequently we are able to rule out the attention-directing hypothesis. Even though participants in the arrows condition were

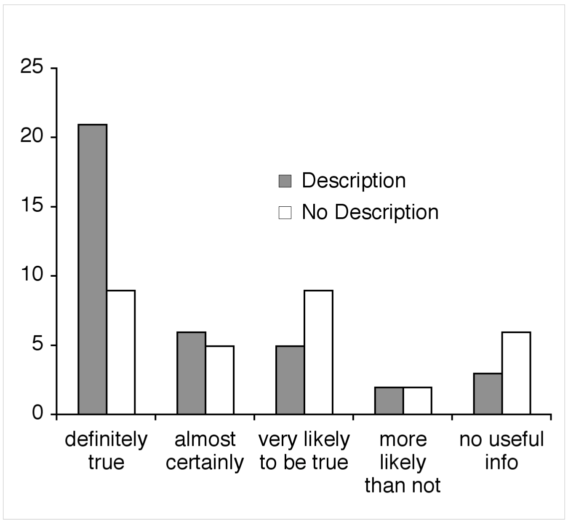

Figure 6. The number of participants selecting each response, from both groups, in Experiment 2.

pointed towards salient parts of the visual evidence, they were less convinced by it than those who saw the descriptive text.

5. Implications.

When discussing the role of visual argumentation in mathematics, authors often refer to the emergence of Cartesian algebraic and analytic methods in the seventeenth century as the point in history when the reputation of visual thinking in mathematics began its decline, and the discovery of false visually obvious statements in the late nineteenth century as the moment when it fell into disrepute. In recent years, visual thinking seems to have regained some of its reputation, mainly because of its huge heuristic potential and because of advances in computer graphics, but it is still regarded by some as simply untrustworthy.

In earlier work we found that mathematicians' reported levels of persuasion in a visual argument could be boosted if an authority figure backed the argument (Inglis and Mejía-Ramos, 2007). This is, in a sense, unsurprising given that the authority figure's knowledge may be thought as being more trustworthy than the unreliable visual representation. Recognising the author of the picture as a knowledgeable person would make the argument, as a whole, more trustworthy. The findings of the present study are more surprising: they suggest that mathematicians tend to be more persuaded by a picture if it comes accompanied by a passage of descriptive text, even if the descriptive text does not go beyond restating what is already clearly pointed out in the picture. Indeed, the participants in this study who saw the descriptive text were significantly more likely to regard the evidence as being definitive: for them it removed all doubts about the claim. This seems to suggest that mathematicians in the description condition were less worried about the epistemological status of the picture than their colleagues in the no-description condition, only because they saw the same information presented in the form of text.

One other interesting observation worth making regards the relative similarity between the response patterns of the two groups of participants in the current study. Both the research-active mathematicians, and the successful mathematics undergraduates exhibited similar effects of similar sizes. Nevertheless, as with earlier research (Inglis and Mejía-Ramos, 2007), although they exhibited similar between-condition effects as the research-active mathematicians, the undergraduate students tended to be less convinced by the visual evidence overall. Across conditions in Experiment 1, the research-active mathematicians ranked the evidence as providing more support to the claim than the undergraduate students, $U = 1064, p = 0.001$. Similarly, more researchers rated the evidence as definitive than undergraduates, $\chi^2(1) = 7.07, p = 0.008$. This finding suggests that undergraduate students tend to be more sceptical of the reliability of visual arguments than their teachers, an observation which is consistent with remarks by earlier mathematicians and mathematics education researchers (Dreyfus, 1994; Littlewood, 1953; Vinner, 1989). Studying the ways in which judgements about mathematical arguments change as students become more mathematically sophisticated would appear to be a fruitful avenue for future research.

6. Concluding Remarks.

In this paper we have reported two brief experiments which interrogated one particular factor which influences the level of persuasion that research-active mathematicians and successful mathematics undergraduates are willing to invest in visual mathematical arguments. Across two experiments it was shown that the presence of a paragraph of text which described the visual image – but provided no deductive support to the claim – made the visual evidence significantly more persuasive. In addition, significantly more participants who saw the descriptive text considered that the visual argument provided definitive support to the claim than those who did not see the text.

Although the experiments reported in this paper are rather basic and exploratory in nature, their findings suggest that the psychological study of mathematical behaviour could productively inform current philosophical discussions regarding the role of visualisation in mathematical practices. Although the study discussed in this paper was experimental, many other empirical methods are also likely to provide useful insights into mathematical behaviour. Dunbar and Blanchette (2001), for example, combined experimental and naturalistic enquiry methods during their programme of research on scientific argumentation (Dunbar and Fugelsang, 2005; Fugelsang et al., 2004). Regardless of the exact nature of the methods adopted, we believe more in-depth empirical studies of actual mathematical behaviour will considerably add to our knowledge regarding not only visual thinking in mathematics, but also the more general topic of mathematical argumentation practices.

Appendix

A. The instructions given to participants.

Please read the following instructions carefully.

This study is concerned with the type of evidence which mathematicians and mathematics students find convincing. On the following pages you will be shown a series of mathematical claims, together with some evidence which is related to the claim. Your task is to evaluate the extent to which the given evidence, and only the given evidence, provides support to the claim. For each problem you will be asked to fill in a form like this:

Please tick one of the options.

The given evidence suggests that the claim is:

- ◯ definitely true
- ◯ almost certain to be true
- ◯ very likely to be true
- ◯ more likely to be true than not; OR
- ◯ this gives no useful information about the claim.

To be clear, we are interested in how you think the given evidence justifies the truth of the claim: not in the truth of the claim itself. So, if you know the claim to be true, but believe the given evidence does not itself offer any support to the claim, you should tick “this gives no useful information about the claim”.

The experiment consists of this instructions page, and six problems. Please work through the problems in order. Please do not return to a problem once you have finished and moved on to another.

Thank you very much for your help. Please click “next” to continue.

Authors' Vitae

Matthew Inglis
Matthew Inglis is a lecturer in the Mathematics Education Centre at Loughborough University. Before joining Loughborough he was a research fellow in the Learning Sciences Research Institute at the University of Nottingham. His earlier studies were at the University of Warwick, where he graduated with a BSc in mathematics and a PhD in mathematics education. His research is focussed on characterising the reasoning behaviour of expert mathematicians, with a specific interest in the relationship between conditional reasoning and the study of advanced mathematics. In addition, he also maintains an active interest in cognitive psychology, and has conducted several studies of visual search behaviour and numerical cognition.

Juan Pablo Mejía-Ramos
Juan Pablo Mejía-Ramos is a postgraduate research fellow at the University of Warwick, where he is currently completing a PhD in mathematics education. Before enrolling as a doctoral student Juan Pablo completed a 4-year undergraduate program in mathematics at Universidad de Los Andes (Bogotá, Colombia) and obtained an MSc degree in mathematics education also from the University of Warwick. In 2009, he will join the Graduate School of Education and the Mathematics Department of Rutgers University as an Assistant Professor. Juan Pablo is mainly interested in mathematical reasoning, particularly the ways in which students construct, read and present arguments in mathematics, and how these argumentative activities develop into those of research-active mathematicians.

Acknowledgements

We thank James Adelman and two anonymous reviewers for their valuable comments and suggestions.

References

Alcock, L. and K. Weber: 2005, 'Proof validation in real analysis: Inferring and checking warrants'. *Journal of Mathematical Behavior* **24**, 125–134.

Appiah, K. A.: 2008, *Experiments in ethics*. Cambridge, MA: Harvard University Press.

Barwise, J. and J. Etchemendy: 1991, 'Visual information and valid reasoning'. In: W. Zimmerman and S. Cunningham (eds.): *Visualisation in teaching and learning mathematics*. Washington, DC: MAA, pp. 9–24.

Brown, J. R.: 1999, *Philosophy of mathematics: as introduction to the world of proofs and pictures*. New York: Routledge.

Corfield, D.: 2003, *Towards a Philosophy of Real Mathematics*. Cambridge: Cambridge University Press.

Dove, I.: 2002, 'Can pictures prove?'. *Logique et Analyse* **45**(179-180), 309–340.

Dreyfus, T.: 1994, 'Imagery and Reasoning in Mathematics and Mathematics Education'. In: D. F. Robitalle, D. H. Wheeler, and C. Kieran (eds.): *Selected Lectures from the 7th International Congress on Mathematical Education*. Quebec, Canada: Les Presses de l'Université Laval, pp. 107–122.

Duffin, J. and A. Simpson: 1993, 'Natural, Conflicting and Alien'. *Journal of Mathematical Behavior* **12**, 313–328.

Dunbar, K. and I. Blanchette: 2001, 'The in vivo / in vitro approach to cognition: The case of analogy'. *Trends in Cognitive Science* **5**, 334–339.

Dunbar, K. and J. Fugelsang: 2005, 'Scientific thinking and reasoning'. In: K. J. Holyoak and R. Morrison (eds.): *Cambridge Handbook of Thinking and Reasoning*. Cambridge: Cambridge University Press, pp. 705–726.

Fugelsang, J., C. Stein, A. Green, and K. Dunbar: 2004, 'Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory'. *Canadian Journal of Experimental Psychology* **58**, 132–141.

Giaquinto, M.: 2007, *Visual thinking in mathematics*. Oxford: Oxford University Press.

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

Hanna, G. and N. Sidoli: 2007, 'Visualisation and proof: a brief survey of philosophical perspectives'. *ZDM Mathematics Education* **39**, 73–78.

Harel, G.: 2007, 'Students' proof schemes revisited: historical and epistemological considerations'. In: P. Boero (ed.): *Theorems in school: From history, epistemology and cognition to classroom practice*. Rotterdam: Sense, pp. 65–78.

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

Harel, G. and L. Sowder: 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. and C. Hoyles: 2000, 'A Study of Proof Conceptions in Algebra'. *Journal for Research in Mathematics Education* **31**(4), 396–428.

Inglis, M. and J. P. Mejía-Ramos: 2007, 'The authoritarian proof scheme'. Presented at the 10th Conference on Research in Undergraduate Mathematics Education, San Diego.

Inglis, M. and J. P. Mejía-Ramos: 2008, 'How persuaded are you? A typology of responses'. *Research in Mathematics Education* **10**.

Inglis, M., J. P. Mejía-Ramos, and A. Simpson: 2007, 'Modelling mathematical argumentation: The importance of qualification'. *Educational Studies in Mathematics* **66**, 3–21.

Klauer, K. C., C. Stahl, and E. Erdfelder: 2007, 'The abstract selection task: New data and an almost comprehensive model'. *Journal of Experimental Psychology: Learning, Memory and Cognition* **33**, 680–703.

Littlewood, J. E.: 1953, *A mathematician's miscellany*. London: Methuen.

Mancosu, P., K. F. Jorgensen, and S. A. Pedersen (eds.): 2005, *Visualization, explanation and reasoning styles in mathematics*. Dordrecht, The Netherlands: Springer.

Martin, W. G. and G. Harel: 1989, 'Proof frames of preservice elementary teachers'. *Journal for Research in Mathematics Education* **20**, 41–51.

Nadelhoffer, T. and E. Nahmias: 2007, 'The past and future of experimental philosophy'. *Philosophical Explorations* **10**, 123–149.
Pinto, M. and D. O. Tall: 2002, 'Building formal mathematics on visual imagery: A case study and a theory'. *For the Learning of Mathematics* **22**, 2–10.
Raman, M.: 2002, 'Proof and justification in collegiate calculus'. Ph.D. thesis, University of California, Berkeley.
Raman, M.: 2003, 'Key ideas: What are they and how can they help us understand how people view proof?'. *Educational Studies in Mathematics* **52**, 319–325.
Recio, A. and J. Godino: 2001, 'Institutional and Personal Meanings of Mathematical Proof'. *Educational Studies in Mathematics* **48**, 83–99.
Reips, U.-D.: 2000, 'The Web Experiment Method: Advantages, Disadvantages, and Solutions'. In: M. H. Birnbaum (ed.): *Psychological Experiments on the Internet*. San Diego: Academic Press, pp. 89–117.
Reips, U.-D. and C. Neuhaus: 2002, 'WEXTOR: A web-based tool for generating and visualising experimental designs and procedures'. *Behavior Research Methods, Instruments and Computers* **34**, 234–240.
Reips, U.-D. and S. Stieger: 2004, 'Scientific Loganalzyer: A web-based tool for analyses of server log files in psychological research'. *Behavior Research Methods, Instruments and Computers* **36**, 304–311.
Segal, J.: 2000, 'Learning about mathematical proof: conviction and validity'. *Journal of Mathematical Behavior* **18**(2), 191–210.
Selden, A. and J. Selden: 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.
Toulmin, S.: 1958, *The uses of argument*. Cambridge: CUP.
Van Kerkhove, B. and J. P. van Bendegem (eds.): 2006, *Perspectives on Mathematical Practices: Bringing Together Philosophy of Mathematics, Sociology of Mathematics, and Mathematics Education*. Dordrecht: Springer.
Vinner, S.: 1989, 'The avoidance of visual considerations in calculus students'. *Focus on Learning Problems in Mathematics* **11**, 149–155.
Weber, K.: 2008, 'How do mathematicians determine if an argument is a valid proof?'. *Journal for Research in Mathematics Education* **(to appear)**.
Weber, K. and L. Alcock: 2004, 'Semantic and syntactic proof productions'. *Educational Studies in Mathematics* **56**, 209–234.

Address for Offprints:
Matthew Inglis
Mathematics Education Research Centre
Loughborough University
Loughborough, United Kingdom.
e-mail: matthew.inglis@nottingham.ac.uk