Loughborough University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

How persuaded are you? A typology of responses

Matthew Inglis[a] and Juan Pablo Mejia-Ramos[b]

[a]*Learning Sciences Research Institute, University of Nottingham, Nottingham, UK;*

[b]*Graduate School of Education, Rutgers University, New Brunswick, NJ, USA*

Several recent studies have suggested that there are two different ways in which a person can proceed when assessing the persuasiveness of a mathematical argument: by evaluating whether it is personally convincing, or by evaluating whether or not it is publicly acceptable. In this paper, using Toulmin's (1958) argumentation scheme, we produce a more detailed theoretical classification of the ways in which participants can interpret a request to assess the persuasiveness of an argument. We suggest that there are (at least) five ways in which such a question can be interpreted. The classification is illustrated with data from a study that asked undergraduate students and research-active mathematicians to rate how persuasive they found a given argument. We conclude by arguing that researchers interested in mathematical conviction and proof validation need to be aware of the different ways in which participants can interpret questions about the persuasiveness of arguments, and that they must carefully control for these variations during their studies.

Keywords: arguments; persuasion; proof.

The issue of what types of arguments students find persuasive and convincing has been a recurring theme in the mathematics education literature. Mason, Burton, and Stacey (1982) placed conviction at the heart of mathematical practice by proposing that learners first need to convince themselves, then a friend, and finally an enemy. Harel and Sowder (1998) analysed the notion of conviction in depth, by publishing a taxonomy of 'proof schemes': the types of arguments that students use both to ascertain for themselves (to remove their own doubts) and to persuade others (to remove others' doubts) about the truth of a statement. Other authors have suggested that these two processes may be disassociated: that the types of arguments that students find personally persuasive may not necessarily be the same types that they would use to persuade third parties such as their mathematics teachers or lecturers (Healy and Hoyles 2000; Mejia-Ramos and Tall 2005; Segal 2000; Raman 2002). In this paper we suggest that the notion of persuasion and conviction is yet more complicated than this two-way categorisation suggests. Here, we argue that there are (at least) five different interpretations that a participant may reasonably make when asked how persuasive they find a mathematical argument.

Toulmin's argumentation scheme

To situate our proposed taxonomy of the ways in which participants may respond to questioning about how 'persuaded' or 'convinced' they are by a mathematical argument, we first introduce Toulmin's (1958) argumentation scheme. Toulmin advocated an approach to analysing arguments that departed dramatically from traditional approaches to formal logic. He was less concerned with the logical validity of an argument, and more interested in the semantic content and structure in which it fits. This manner of analysing argumentation has become known as 'informal logic' in order to emphasise its differences from formal logic.

Toulmin's scheme has six basic types of statement, each of which plays a different role in an argument. The *conclusion* (C) is the statement of which the arguer wishes to convince their audience. The *data* (D) are the foundations on which the argument is based, the relevant evidence for the claim. The *warrant* (W) justifies the connection between data and conclusion by, for example, appealing to a rule, a definition, or by making an analogy[1]. The warrant is supported by the *backing* (B), which presents further evidence. The *modal qualifier* (Q, henceforth *qualifier*) qualifies the conclusion by expressing degrees of confidence; and the *rebuttal* (R) potentially refutes the conclusion by stating the conditions under which it would not hold. Importantly, in any given argument, not all of these statements will necessarily be explicitly verbalised. These six components of an argument are linked together in the structure shown in Figure 1.

Insert Figure 1 about here

In the field of mathematics education, many researchers have applied Toulmin's scheme to analyse arguments constructed by students. However, it has become commonplace to use a reduced version of the scheme by omitting the qualifier and rebuttal. Krummheuer, for example, adopted this reduced version of the scheme to analyse pupil behaviour throughout his long programme of research on the development of collective argumentation practices in primary school classroooms (e.g. Krummheuer 1995, 2007). A similar stance has been adopted by researchers studying classroom interaction at the university level (Stephan and Rasmussen 2002), basic number skills (Evens and Houssart 2004), logical deduction (Hoyles and Küchemann 2002; Weber and Alcock 2005), geometry (Cabassut 2005; Pedemonte 2005), and general proof (Yackel 2001). Inglis, Mejia-Ramos and Simpson (2007) argued that without using Toulmin's full scheme it may be difficult to model

accurately the full range of mathematical argumentation. They gave research-active mathematicians a series of conjectures, and asked them to decide whether or not the conjectures were true, and to provide proofs. It was found that these mathematicians regularly constructed arguments with non-deductive warrants in order to *reduce* rather than *remove* their doubts about a conjecture's truth value. Inglis et al. (2007) pointed out that it would be impossible to model such arguments accurately without incorporating the qualifier component of Toulmin's scheme. They concluded that (i) using the restricted version of Toulmin's scheme in the manner adopted by earlier researchers reduces the range of mathematical arguments that can be successfully modelled; and (ii) rather than concentrating on the appropriateness of the warrants deployed by students, researchers should instead study the appropriateness of the warrant-qualifier pairings constructed in student argumentation.

In this paper, we use Toulmin's full scheme to derive a classification of the ways in which the question "how persuaded are you by this argument?" can be interpreted. We note that, as in the case of argument construction, a comprehensive study of argument evaluation cannot be conducted using a reduced version of the scheme.

How persuaded are you?

Some previous researchers have studied the types and levels of conviction and persuasion students place in an argument simply by asking them (e.g. Mejia-Ramos and Tall 2005; Segal 2000; Raman 2002). But what do students understand by such a question? We suggest that there are (at least) five distinct and reasonable ways of answering the question *how persuaded are you by this argument?* To demonstrate our typology in a general context, we introduce the following fictional day-to-day argument about train times:

> The last three times I have gone anywhere by train, I have arrived several hours late. So, it is
>
> certain that, when I go to the airport tomorrow, the train will be late.

Insert Figure 2 about here

Using Toulmin's (1958) layout we can model this argument by identifying its different types of statements (see Figure 2). We claim that a person evaluating how persuaded he or she is by this argument may focus on different parts and aspects of the argument. He or she may focus on: (i) the *data* of the given argument, and how significant/trustworthy it is; (ii) the likelihood of its *conclusion*; (iii) the strength of the *warrant* (and its associated *backing*); (iv) the given *qualifier* (and its associated *rebuttal*), and the extent to which this qualifier is appropriate considering the rest of the argument; and (v) the particular *context* in which the given argument may take place. We now focus on each of these evaluation types in turn.

Type 0

One possible evaluation occurs when the participant focuses on the *data* of the given argument and evaluates the whole argument in terms of how strongly he or she trusts these data.

One possibility is that the participant distrusts the data and projects these doubts onto his/her evaluation of the whole argument. For instance, in the train example someone may suspect that the arguer is lying about the irregularities of the three train trips mentioned. For that person, this argument could be rated as unpersuasive mainly because he or she considers its data unreliable.

On the other hand, it is possible that someone's feeling of affinity for the data would be so strong that he or she would feel persuaded by the argument without taking into account

how these data fit in the whole argument. For example, someone who has herself arrived

several hours late in her last few train trips may find the argument persuasive mainly because

of her empathy for its data.[2] Such an evaluation may involve what is known in the

psychology literature as *myside bias,* a tendency to evaluate data from one's own perspective,

having difficulty in decoupling one's prior beliefs and opinions from the evaluation of

evidence and arguments (Stanovich and West 2006). In this particular case, the evaluator

could focus on the data and evaluate them in a manner biased towards his or her own

opinions. These opinions and beliefs could then be projected on to the evaluation of the

whole argument.

Type 1

Another way of evaluating this argument is by focusing on the likelihood of its *conclusion*

alone. In the train example, someone focusing on the argument's conclusion (i.e. "the train to

the airport will be late tomorrow") may report being highly persuaded by the argument, since

he or she knows that scheduled track repairs in the vicinity of the airport will indeed delay

trains that day. Alternatively, some people may not feel persuaded by the argument as,

knowing that the train to the airport is the most reliable journey in the local train network,

they expect this train to be on time.

In these two cases, evaluators would have reported the qualifier component of an

entirely separate, and self-constructed, argument to that which they were asked to evaluate;

the only similarities with the original argument being that they shared conclusions. In other

words, in these cases the evaluators have their own evidence and reasons for

trusting/distrusting the conclusion, and this information is projected onto the evaluation of the

whole argument. However, as in Type 0 evaluations, it may also be the case that evaluators'

uninformed intuition regarding the argument's conclusion, and not explicit information

external to the argument, influence their reported level of persuasion.

Type 2

Another possible evaluation occurs when participants focus their attention on the *warrant* of

the argument (with its associated *backing*). Unlike the data and the conclusion, the warrant of

the argument is inextricably linked to other parts of the argument: it is a statement linking the

data and the conclusion. Furthermore, any question regarding the trustworthiness of the

warrant would lead to an evaluator querying its (explicit or implicit) backing and to consider

possible rebuttals: if the warrant is appropriately backed and accepts no rebuttals (or only

extraordinary ones), then one would say that the argument strongly supports the conclusion;

but if the warrant is not satisfactorily backed and one can think of critical rebuttals, one

would say that the warrant weakly supports the conclusion. Therefore, focusing on the

warrant of an argument, a person may evaluate the strength with which it links the data with

the conclusion, taking into account its backing and possible rebuttals. In this case, a

participant's evaluation of the whole argument essentially consists of completing this core

part of the given argument (data-warrant-conclusion) with what he or she believes is the

appropriate qualifier for the argument, and then reporting this qualifier as his or her level of

persuasion in the whole argument.

In the train example, a person might reply to the request of stating his or her level of

persuasion in the whole argument by saying that (given the data, implicit warrant and the

possible rebuttals associated with it) it is reasonable to reach the conclusion with a plausible

qualifier. It is important to note that in this case the person is paying little or no attention to

the absolute qualifier that was actually given in the train argument; he or she would be

reporting what they believed to be the appropriate qualifier. It is also important to note that

this way of evaluating the argument differs from Type 1 evaluations: a Type 2 evaluation focuses on the given warrant and takes into account certain information from the argument that is associated with that warrant, whereas a Type 1 evaluation focuses on its conclusion and may involve the (possibly implicit) construction of an entirely new argument.

Type 3

In contrast to the previous types of evaluation, a participant's attention may be drawn to the *qualifier* given in the argument (and its associated *rebuttal*). A Type 3 evaluation occurs when the evaluator decides to what extent he or she believes that the given qualifier is appropriate, considering the rest of the argument.

In the train example, someone may state that he or she is not at all persuaded by the argument as, although it might be reasonable to be worried about the possible lateness of the train based on prior experience, it is completely inappropriate to pair such a warrant with an absolute qualifier as the arguer appears to have done. Unlike a Type 2 evaluation, where the evaluator decides what type of qualifier would be appropriate given the rest of the argument, in a Type 3 evaluation the issue is whether *the given qualifier* is appropriate on account of the rest of the argument. It is clear that one could simultaneously consider an argument to be Type 2 persuasive but Type 3 unpersuasive. Indeed, believing (based on the prior experience cited in the argument) that the train would plausibly – but not certainly – be late could lead someone to such a judgement: they would assess the *appropriate* qualifier to be relatively high (i.e. Type 2 persuasive), but the qualifier *as given in the argument* to be inappropriate (i.e. Type 3 unpersuasive).

Type 4

Finally, instead of focusing on a particular part of the argument, the participant may attend to the *context* in which the argument is situated, and the kinds of arguments that are admissible in such contexts. In this case, when asked how persuaded they are by a given argument, participants may answer by considering how acceptable the argument would be in a particular context. It is well known in the context of jurisprudence that some arguments, no matter how persuasive, are not admissible in court. In England and Wales, for example, a prosecuting lawyer may not refer to a defendant's criminal record during the case. An argument based on such data may well carry an extremely high qualifier, but in the given context it is inadmissible. Naturally what constitutes an admissible argument will depend on the particular context: what is admissible in a criminal court is different from what is admissible in a civil court which, in turn, is different from what is admissible during an argument in a pub.

The example of the train argument may well be admissible when talking informally, but if one were attempting to convince one's departmental finance officer to issue an advance to pay for a taxi fare to the airport, it could be considered inadmissible. Such matters are governed by a set of rules which state what kinds of data, warrants, backings, qualifiers and rebuttals can be used in an admissible argument; and a hunch about the possible lateness of the train is unlikely to meet these rules.

The five types of persuasion we have discussed are summarised in Table 1.

Insert Table 1 about here

Illustrating the typology in mathematics

Our primary aim in the second half of the paper is to illustrate the applicability of this typology to the evaluation of mathematical arguments. A second aim is to provide 'existence proofs' of each of the types: to show that the different types of evaluations can be, and are,

made by mathematicians and students when evaluating mathematical arguments (or, at least, that mathematicians and students claim to be making evaluations of each type). To accomplish these aims, we draw on evaluations of a heuristic argument collected recently as part of a study on the role of authority in mathematical argumentation (Inglis and Mejia-Ramos 2006). The argument used in the study was given by Gowers (2006) and supports the conjecture that there are one million consecutive sevens somewhere in the decimal expansion of π:

> 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. (p. 194).

The two stages of this argument are shown graphically, using Toulmin's (1958) scheme, in Figure 3.

Insert Figure 3 about here

Our sample consisted of two groups, undergraduate students and research-active

mathematicians. Participants completed the task online (for a discussion on the reliability of

internet studies see, for example, Krantz and Dalal 2000). The undergraduate students were

studying at one of four highly-regarded UK universities, and were asked to participate by

means of an email from their 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 were recruited in

a similar manner as the undergraduates, via emails from their departmental secretaries; others

were recruited through an advertisement posted on a mathematics research newsgroup. In the

study, participants were presented with Gowers' (2006) argument, and were asked to state to

what extent they were persuaded by it, using either a five point Likert scale or a continuous

0–100 scale (depending on whether participants took part in the pilot or main study). In

addition, participants were invited to leave explanatory comments on their reported level of

persuasion. It is these comments that we use in the following sections to illustrate our

theoretical classification.

Our focus in the analysis of the extracts is on the aspects of the argument which

participants focus upon when explaining their evaluations. While we accept that the full

complexity of participants' judgements may not be fully reflected by such short explanations

(especially given factors such as *myside bias*, Stanovich and West 2006), we nevertheless

believe that focussing on these reported comments will allow us to illustrate the utility of the

typology for researchers interested in how mathematicians and students become persuaded by mathematical arguments.

We should emphasise that the classification we introduce in this paper is derived from a *theoretical* analysis of Toulmin's (1958) argumentation scheme; an analysis that considers the type of statement upon which participants may focus their attention when asked to evaluate their degree of persuasion in an argument. The extracts reported in the following sections, therefore, should be viewed as existence proofs of each of the categories, not as a dataset from which we are attempting to generalise.

Type 0

When asked to evaluate and explain their level of persuasion in Gowers' argument regarding the decimal expansion of π, one research-active mathematician wrote:

> This is not an argument. To be more precise, in the statement no concrete evidence is presented. He only explains how the statistical evidence could look like, but does not specify the empirical results of the tests. (Researcher).

In this case, the evaluator's explanation of his or her rating clearly focused on the data of the first stage of the argument. This researcher's comment concentrated on the "statistical evidence" of the argument, reporting dissatisfaction with the lack of concreteness of its presentation. Although this factor may be related to concerns that the researcher may have had with the warrant and backing of the argument, it is clear that their focus was on the data: to persuade this participant, at a minimum the data of the argument would need to be presented in a considerably more formal fashion.

Type 1

Type 1 evaluations were not uncommon among participants' comments. For example, one research-active mathematician wrote:

Normalcy of (the digits) of pi is not unreasonable given almost all reals are normal.

(Researcher).

This researcher's comment focused on the conclusion of the first stage of the argument (i.e.

"π is normal"), and reported the qualifier component (i.e. "not unreasonable") of an entirely

separate, and self-constructed, argument to that which participants were asked to evaluate. He

or she did not mention the data, warrant, backing or rebuttal given in Gowers' argument, and

instead appears to have constructed a separate argument that merely shares a conclusion with

the given argument. The data ("almost all reals are normal"), implicit warrant, and implicit

backing in this new argument are entirely distinct, and the evaluator reported the new

argument's qualifier by stating that the conclusion is "not unreasonable".

Another example of a Type 1 evaluation came from the following student:

I am mainly not persuaded because I have seen a formula which can calculate the n-th digit of

pi, suggesting that is not a random series of numbers. (Undergraduate Student).

Again, the student's comment focused on the conclusion of the first stage of Gowers'

argument, and evaluated the qualifier of an entirely new argument; the data, warrant and

backing of the given argument are not taken into account. Of course, the construction of an

entirely new argument – necessary for a Type 1 evaluation – would only be possible if the

participant had a strong background knowledge of the domain in which the argument is

situated.

Type 2

The following response typifies a Type 2 evaluation of Gowers' argument:

The evidence lends decent weight to the conjecture; but naturally as proof is impossible it is

unrealistic to assume certainty. (Undergraduate Student).

Here the student seemed to be suggesting that the evidence presented – the data, warrant and

backing – indicates that the conclusion may be true, but that any stronger qualifier would be

inappropriate (possibly considering the existence of rebuttals). This student has considered the warrant of the argument (with its associated data, backing, and possible rebuttals) and has decided that he or she would be willing to pair a qualifier with it that "lends decent weight" to the conclusion. Characteristically for Type 2 evaluators, this student does not seem to be addressing the argument's *given qualifier*; this comment only refers to the new qualifiers that he or she considers appropriate given the rest of the argument.

Type 3

When participants were asked to evaluate Gowers' argument there were many examples of Type 3 evaluations:

> Despite the statistical evidence that pi is a 'normal number' (only testing short sequences) there could still be some subtle numerical invariant that prevents this particular very long sequence from occurring. (Researcher).

In this comment, the researcher first evaluated the link between the data of the first stage of the argument and its conclusion, and then centred on an unmentioned rebuttal (the possibility of a "subtle numerical invariant"), suggesting that Gowers' given qualifier was inappropriate. A similar evaluation was reported in the following comment by an undergraduate student:

> The reasoning is flawed in moving from talking of experience strongly suggesting 'short sequences' occur in naturally occurring irrational numbers to saying that 'a million sevens' is likely to occur. Of course, their definition of a 'short sequence' isn't given, but I dare guess it is much fewer numbers than a million. (Undergraduate Student).

Again, in this case the student criticised the strength with which the data *is claimed* to support the argument's conclusion, focusing on what he or she considered to be an inappropriate qualifier. In this comment, the student did not report the extent to which he or she believed that the given evidence supports the conclusion; instead they seemed to be more concerned about the relatively high qualifier given in the original argument.

In both these extracts the evaluators exhibited the hallmarks of a Type 3 evaluation; they explained that they were not persuaded by the argument as a whole because they did not accept that the given warrant (and associated backing) justifies the given qualifier (and associated rebuttal).

Type 4

A Type 4 evaluation can clearly be seen in this researcher's response:

> The argument hinges on a precise notion of randomness in the digits of pi, which may be plausible, but hasn't been proven. If a manuscript that made an analogous argument came to me for refereeing, I'd recommend it be rejected for lack of mathematical rigour. However, if someone wanted to generate 'good pseudorandom' bits from the digits of pi for a casual computer program (i.e., not one on which lives or property crucially depend), I'd say Gowers' argument would justify the strategy. (Researcher)

Here the evaluator suggested that in the context of an academic mathematics journal he or she would deem the argument to be inadmissible, but that in a different context, where one merely needed to generate some random numbers, it would be admissible. This position was clarified still further by noting that in yet another context – where the random numbers were a matter of life or death – then perhaps the argument would again struggle to meet the requirements of admissibility.

Within the context of mathematics education Type 4 evaluations are very important, and an ability to understand successfully the different rules of admissibility for different contexts may be a hard skill for students to develop. These rules undoubtedly vary between educational levels – the type of justification required at school level mathematics is typically very different from that required at university level – but may also vary between courses at a single level. For example, the types of argument which are admissible to justify the rules of integration may be very different if the notion of integral is studied during a real analysis

course compared with during an applied fluid dynamics course. In the former a formal derivation from the definition of, for example, the Riemann integral, may be necessary, whereas in the latter a statement of the result might be sufficient.

Mixing the types

Sometimes participants in empirical research studies may give evaluations of different types in the same response: multiple interpretations of the question may lead to answers with multiple layers. This, for example, is how one mathematics researcher responded when asked to evaluate Gowers' (2006) argument:

> Purely logically on the basis of the evidence presented, I am not persuaded at all. However, I am aware that there is a substantial body of research (rather more formal than the waffle above) specifically addressing equidistribution of digit sequences of pi. So I moved from the most sceptical to the next category by way of combining that knowledge with the information above. (Researcher)

Here the evaluator explicitly noted that he or she was not persuaded by the data, warrant and backing of the argument: they were completely Type 2 unpersuaded. However, the researcher claimed that they were somewhat Type 1 persuaded on account of his or her background knowledge about the digit sequences of π. Understandably, these differing interpretations gave this participant some difficulty when asked to rate his or her level of persuasion on a Likert scale. However, the proposed typology can help us to make sense of this participant's multi-layered written comment.

One question, five ways of answering

Earlier researchers have studied two different ways of evaluating a given argument and the two corresponding levels of persuasiveness reported by their participants. This led them to establish a distinction between a private and a public, or internal and external, sense of

conviction. Raman (2003, 320), for example, differentiated between private and public

arguments and their corresponding senses of conviction (see also Raman 2002):

> By 'private argument' I mean, 'an argument which engenders understanding', and by 'public'
>
> I mean 'an argument with sufficient rigor for a particular mathematical community'.

In our terms, Raman was noting the differences between Type 4 persuasion and persuasion of

Types 0–3. The usefulness of this two-way distinction in mathematics education lies in the

importance of Type 4 evaluations in mathematical argumentation. Mathematical proof seems

to set argument admissibility in mathematical practice aside from admissibility standards in

other contexts, making students' beliefs of what constitutes a valid mathematical proof, and

the ways in which these beliefs influence their reported level of persuasion in a given

argument, an interesting topic of study among mathematics educators.

However, we suggest that a finer typology of persuasiveness may be helpful: whereas

earlier researchers have spoken only of a 'private' sense of conviction, we have demonstrated

that, in the case of argument validation, there are (at least) four different ways in which such

an evaluation may be conducted in a 'private' fashion. Similarly, there is not simply one

variety of 'public', or Type 4, persuasion. Each particular context brings with it its own

particular rules of admissibility, and these rules vary greatly between contexts. Even in the

particular context of mathematics, requirements for rigour alter greatly according to

educational level, mathematical subject and other particular circumstances of each

evaluation.

Segal (2000) used two different questions to study this distinction between a private

and public sense of conviction. Following Mason et al. (1982), she asked her participants

whether or not a given argument convinced them personally, and whether or not the argument

would persuade "one's enemies (as opposed to one's friends, or oneself)" (Segal 2000, 199).

It is unclear whether a participant's response to the first question involves an evaluation of

Type 0, 1, 2 or 3. Furthermore, a participant's response to the second question (arguably related to Type 4 evaluations) is, of course, dependant upon in which context the participant situates their enemy, and what type of evaluation is expected from that enemy.

Therefore, this finer typology may be used by both teachers and researchers not only to better assess students' and participants' reported levels of persuasion in a given mathematical argument, but also to design specific questioning strategies to incline students and participants towards making the teacher's or researcher's desired type of evaluation.

Using the typology: an example from the literature

In this section we give a specific example of a piece of analysis from the mathematics education literature to illustrate the utility of the typology proposed here. During their study of the proof conceptions of school children, Coe and Ruthven (1994) looked at students' investigative and problem solving strategies. Here we concentrate on one particular extract from an interview transcript reported by Coe and Ruthven. Whilst working on a problem regarding the sums of diagonals in a number square, Bill, a 17 year old student, checked that a statement he was investigating was true for six cases, and then said that it was "safe to make a conjecture". The interviewer pressed him by asking "what sort of percentage certainty would you put behind that, say, if I forced you on that?" Bill replied by estimating he had a "percentage certainty" in the "high nineties".

What has happened here? Bill and the interviewer were discussing the persuasiveness of an argument with an inductive warrant (which consisted of the numerical evaluation of six examples). The interviewer pressed Bill with an ambiguous question, by asking the "percentage certainty" he would be willing to "put behind that". One interpretation is that, when pressed, Bill conducted a Type 2 evaluation. He evaluated what sort of qualifier he was willing to pair with the data, warrant and conclusion of the given (self-constructed) argument,

and made the decision that he was willing to deploy a high (but non-absolute) qualifier. (Seen in this light, Bill's behaviour closely matches that of the highly successful research students interviewed by Inglis et al. 2007). Coe and Ruthven (1994), however, appear to have interpreted Bill's response in a different way, as a Type 4 evaluation. They wrote that Bill's "certainty appears to be gained just by checking a relatively small number of cases" (50), and used the episode as evidence for the claim that "students' *proof* strategies were primarily and predominantly empirical" (52, our emphasis). Of course, it may well be that Bill *was* conducting a Type 4 evaluation of his self-constructed argument: there are certainly many studies which corroborate Coe and Ruthven's finding that students often think empirical evidence can form admissible proofs (e.g. Balacheff 1987, Harel and Sowder 1998). However, when seen within the typology set out in this paper, Coe and Ruthven's interpretation of this interview evidence is, at best, arguable.

We suggest that an awareness of the typology presented in this paper could help researchers conducting studies on mathematical conviction to deploy careful questioning strategies to increase the likelihood of accurately interpreting their interviewee's behaviour.

Concluding Remarks

Researchers interested in assessment have, for some time, been aware that there may be a gap between test designers' interpretation of a given question and the interpretation of those who respond to the question. In their study of 11 and 12 year old children's responses to national test items, Cooper and Harries (2002) found that students interpretations of how much realism to use in their answers when answering 'realistic' mathematical questions differ from those of the questions' designers. We suggest that teachers and researchers who are interested in what types of mathematical argument students find persuasive need to have awareness of the differing ways in which their questions may be interpreted. Similarly, the many

researchers who have studied the manner in which students and mathematicians validate proofs (e.g. Selden and Selden 2003; Weber, forthcoming) also need to be aware that requests to 'evaluate' a purported proof may lead to differing interpretations from different participants.

In this paper we have proposed that there are (at least) five different ways in which participants in research studies can reasonably interpret a request to evaluate their level of conviction or persuasion in an argument. The first two types that we have described revolve around a participant evaluating a particular part of the argument (data or conclusion) and paying little or no attention to the other parts; two other types involve the participant evaluating the core part of the argument and either completing it with what they believe is an appropriate qualifier, or assessing whether or not the given qualifier is appropriate; a fifth type is related to one particular context in which the argument may take place and the participant's evaluation of whether or not the given argument would be admissible in such context. By using Toulmin's full scheme, it is possible to distinguish clearly between these different types of argument evaluation. Given these different ways, we suggest that the empirical researcher must design their methodological instruments carefully to determine which question participants are responding to, and take into account these different types of evaluations in their theorisation of students' reported levels of persuasion.

Acknowledgements

Notes

1. Toulmin's (1958) use of the word 'warrant' is not identical to the way in which the term has been used by some mathematics education literature. Rodd (2000), for example, saw a warrant as removing uncertainty, whereas Toulmin was more flexible, accepting that a warrant can be qualified so as to merely reduce uncertainty.

2. Elena Nardi has pointed out to us that this may be the type of persuasion enlisted by propagandists who use sensationalist data in the hope that the affinity of the audience for these data will sway them to support whatever conclusion he or she wishes to draw.

References

Balacheff, N. 1987. Processus de preuves et situations de validation. *Educational Studies in Mathematics*, 18: 147–176.

Cabassut, R. 2005. Argumentation and proof in examples taken from French and German textbooks. In *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education*. Sant Feliu de Guixols, Spain: ERME.

Coe, R., and K. Ruthven. 1994. Proof practices and constructs of advanced mathematical students. *British Educational Research Journal* 20: 41–53.

Cooper, B., and T. Harries. 2002. Children's responses to contrasting 'realistic' mathematics problems: Just how realistic are children ready to be*? Educational Studies in Mathematics* 49: 1–23.

Evens, H., and J. Houssart. 2004. Categorizing pupils' written answers to a mathematics test question: 'I know but I can't explain'. *Educational Research* 46: 269–82.

Gowers, W.T. 2006. Does mathematics need a philosophy? In *18 unconventional essays on the nature of mathematics,* ed. R. Hersh, 182–200. New York: Springer.

Harel, G., and L. Sowder. (1998). Students' proof schemes: Results from exploratory studies. In *Research in Collegiate Mathematics* III*,* ed. A. H. Schoenfeld, J. Kaput, and E. Dubinsky, 234–82. Providence, RI: American Mathematical Society.

Healy, L., and C. Hoyles. 2000. A study of proof conceptions in algebra. *Journal for Research in Mathematics Education* 31: 396–428.

Hoyles, C., and D. Küchemann. 2002. Students' understanding of logical implication. *Educational Studies in Mathematics* 51: 193–223.

Inglis, M., and J.P. Mejia-Ramos. 2006. Is it ever appropriate to judge an argument by its author? In *Proceedings of the British Society for Research into Learning Mathematics,* ed. D. Hewitt, Vol. 26, no. 2, 43–8. University of Bristol: BSRLM.

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

Krantz, J.H., and R. Dalal. 2000. Validity of web-based psychological research. In *Psychological experiments on the internet,* ed. M.H. Birnbaum, 35–60. San Diego: Academic Press.

Krummheuer, G. 1995. The ethnology of argumentation. In *The emergence of mathematical meaning: Interaction in classroom cultures,* ed. P. Cobb and H. Bauersfeld, 229–69. Hillsdale: Erlbaum.

———. 2007. Argumentation and participation in the primary mathematics classroom: Two episodes and related theoretical abductions. *Journal of Mathematical Behavior* 26: 60–82.

Mason, J., L. Burton, and K. Stacey. 1982. *Thinking mathematically*. London: Addison-Wesley.

Mejia-Ramos, J.P., and D. Tall. 2005. Personal and public aspects of formal proof: A theory and a single-case study. In *Proceedings of the Sixth British Congress of Mathematics Education,* ed. D. Hewitt and A. Noyes, 97–104. Coventry: BSRLM.

Pedemonte, B. 2005. Quelques outils pour l'analyse cognitive du rapport entre argumentation et démonstration. *Recherches en Didactique des Mathématiques* 25: 313–348.

Raman, M. 2002. Proof and justification in collegiate calculus. PhD diss., University of California, Berkeley.

———. 2003. Key ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics* 52: 319–25.

Rodd, M.M. 2000. On mathematical warrants: Proof does not always warrant, and a warrant may be other than a proof. *Mathematical Thinking and Learning* 2: 221–44.

Segal, J. 2000. Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior* 18: 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: 4–36.

Stanovich, K.E. and R.F. West. 2006. Natural myside bias is independent of cognitive ability. *Thinking and Reasoning* 13: 225–247.

Stephan, M., and C. Rasmussen. 2002. Classroom mathematical practices in differential equations. *Journal of Mathematical Behavior* 21: 459–90.

Toulmin, S. 1958. *The uses of argument*. Cambridge: CUP.

Weber, K. Forthcoming. How do mathematicians determine if an argument is a valid proof? *Journal for Research in Mathematics Education*.

Weber, K., and L. Alcock. 2005. Using warranted implications to understand and validate proofs. *For the Learning of Mathematics* 25, no. 1: 34–8.

Yackel, E. 2001. Explanation, justification and argumentation in mathematics classrooms. In *Proceedings of the 25th International Conference on the Psychology of Mathematics Education,* ed. M. van den Heuvel-Panhuizen, Vol. 1, 9-23. Utrecht, Holland: IGPME.

Figure Captions

1. The layout of Toulmin's (1958) argumentation scheme, showing data (D), warrant (W), backing (B), qualifier (Q), rebuttal (R) and conclusion (C).

2. Late-train argument modelled using Toulmin's (1958) scheme, with inferred warrant, backing and rebuttal components (the inferred components are italicised).

3. Two stages of Gowers' (2006) argument modelled using Toulmin's (1958) scheme, with inferred components italicised. (A number is said to be *normal* if its digits show a random distribution).

Table Caption

1. A summary of the types of persuasiveness identified in this paper, expressed using Toulmin's (1958) scheme.
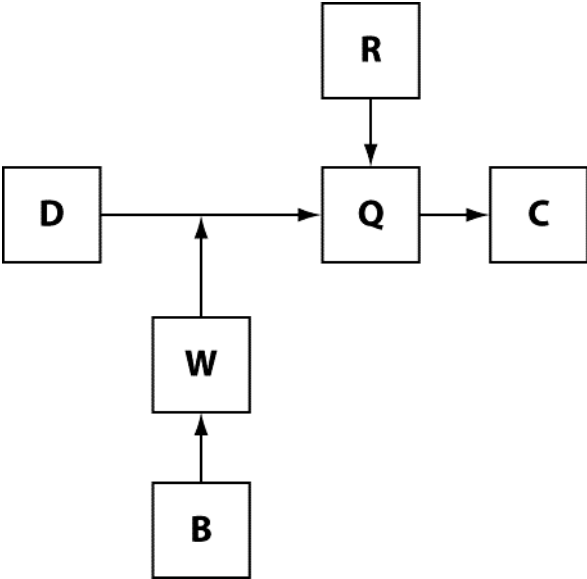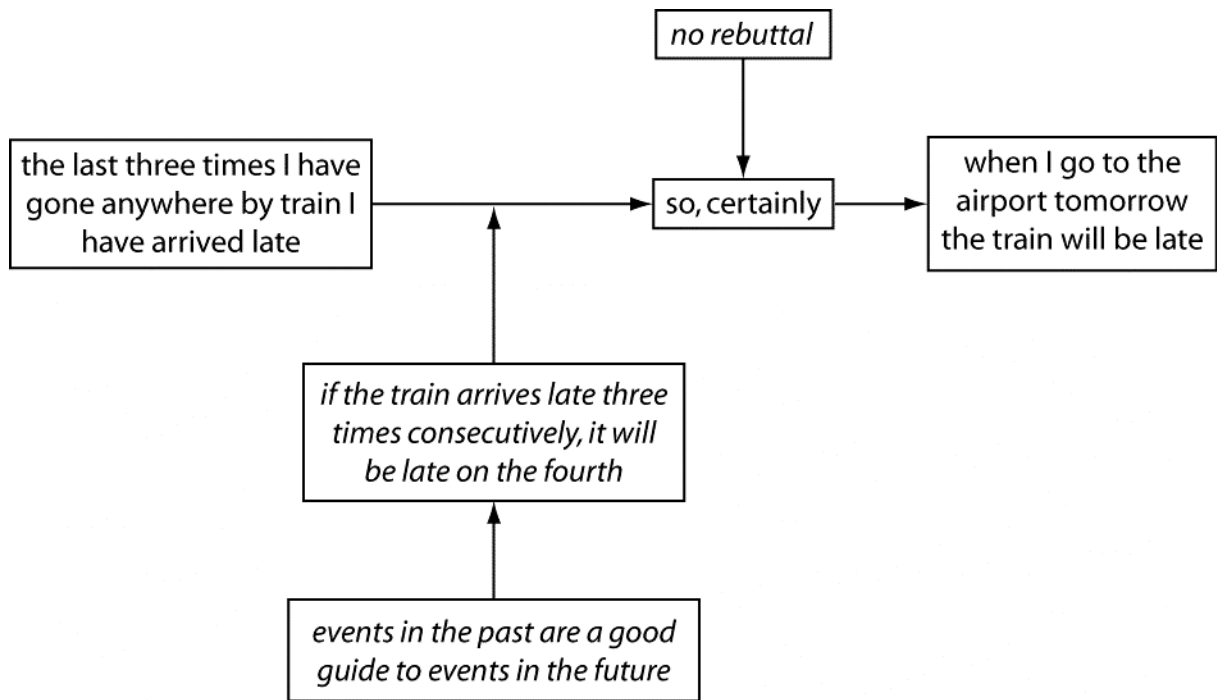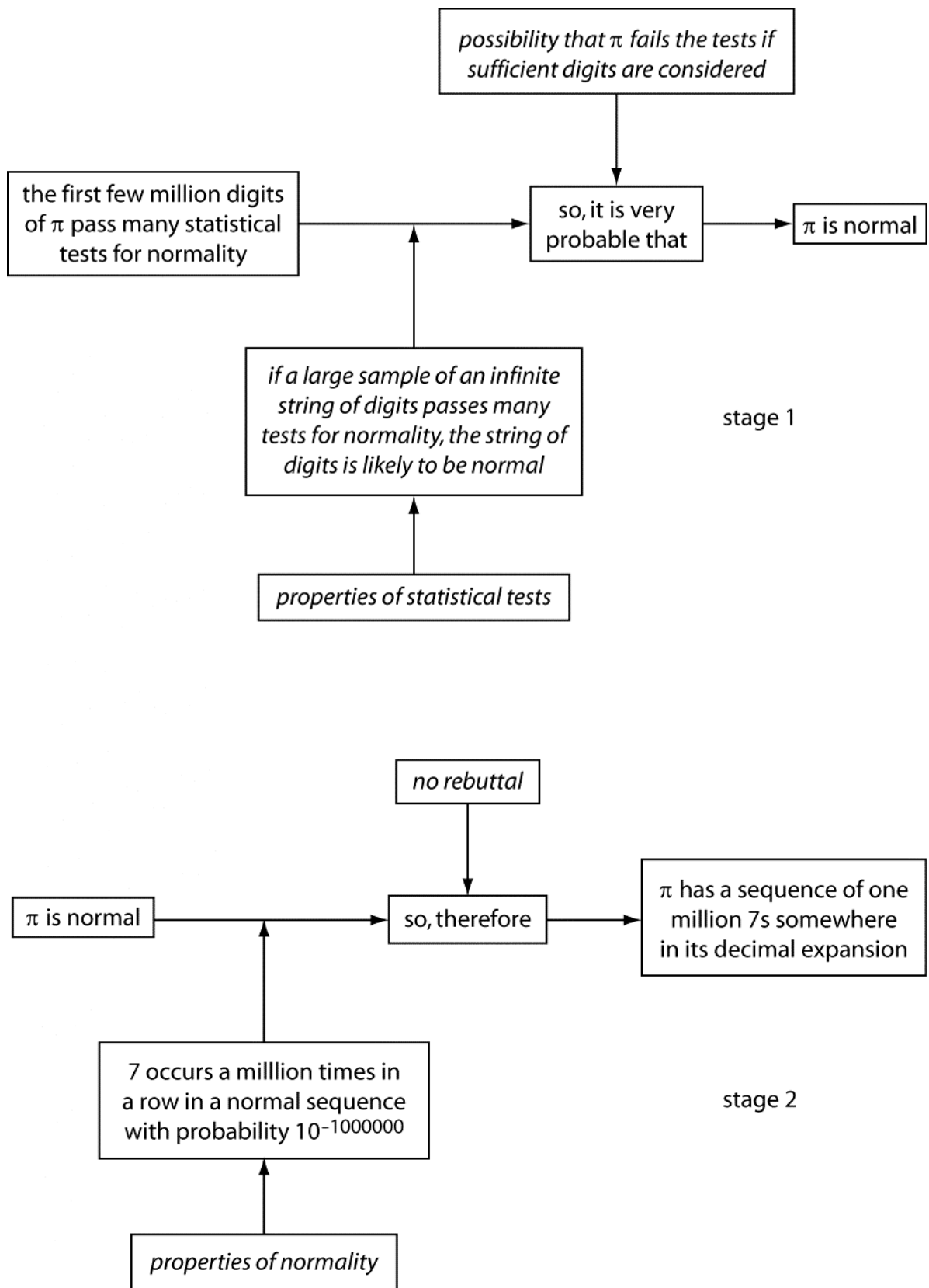
Figure 1.

Figure 2.

Figure 3.

Table 1.

Type	Foci of attention (in **bold**) in the given argument	Question answered in terms of the focus of attention
0.	**Data**	How strongly do you trust the **Data**?
1.	**Conclusion**	How strongly do you believe the **Conclusion**?
2.	Data—**Warrant**—Conclusion | Backing	How strongly does the given **Warrant** (and its associated backing) support the conclusion? (i.e. what is the appropriate qualifier for the argument?)
3.	Rebuttal | Data—Warrant—**Qualifier**—Conclusion | Backing	How appropriate is the **Qualifier** (and its associated rebuttal) given the rest of the argument?
4.	**Context** Rebuttal | Data—Warrant—Qualifier—Conclusion | Backing	How admissible is the argument in a given **Context**?