MATHEMATICIANS AND THE SELECTION TASK

Matthew Inglis & Adrian Simpson

Mathematics Education Research Centre, University of Warwick

Learning to think logically and present ideas in a logical fashion has always been considered a central part of becoming a mathematician. In this paper we compare the performance of three groups: mathematics undergraduates, mathematics staff and history undergraduates (representative of a 'general population'). These groups were asked to solve Wason's selection task, a seemingly straightforward logical problem. Given the assumption that logic plays a major role in mathematics, the results were surprising: less than a third of students and less than half of staff gave the correct answer. Moreover, mathematicians seem to make different mistakes from the most common mistake noted in the literature. The implications of these results for our understanding of mathematical thought are discussed with reference to the role of error checking.

LOGIC IN MATHEMATICS

Learning to think logically appears to be at the heart of almost every university level mathematics course. Stewart & Tall, for example, explain that

> everyday language is full of generalities which are vaguely true in most cases, but perhaps not all. Mathematical proof is made of sterner stuff. No such generalities are allowed: all the statements involved must be clearly true or false [... we must] be sure that our mathematical logic is flawless. (Stewart & Tall 1977, p.110)

The mathematics education literature agrees, Devlin, for example, notes that

> the ability to construct and follow fairly long causal chains [and] a step by step logical argument [...] is fundamental to mathematics. (Devlin 2001, p.15)

Previous work on logic in the mathematics education literature has largely concentrated upon schoolchildren. Hoyles & Küchemann (2002), for example, found that even high achieving Year 9 students often "failed to appreciate how data can properly be used to support a conclusion as to whether $P \Rightarrow Q$ is true or not" (p. 217). This finding mirrors similar results from experiments conducted on the general population (e.g. Oakhill, Johnson-Laird & Garnham 1989). Despite these results, the assumption that the ability to use logic is an essential ingredient in becoming a successful mathematician has remained unchallenged. Perhaps the most famous experiment that demonstrated the lack of logical thought in the general population was conducted by the psychologist Peter Wason.

THE SELECTION TASK

The 'selection task' was first reported by Wason (1968). His experimental setup was elegant and deceptively simple. Participating subjects were shown a selection of cards, each of which had a letter on one side and a number on the other. Four cards were placed on a table:

The participants were given the following instructions:

> Here is a rule: “every card that has a D on one side has a 3 on the other.” Your task is to select all those cards, but only those cards, which you would have to turn over in order to discover whether or not the rule has been violated.

The correct answer is to pick the D card and the 7 card, but across a wide range of published literature only around 10% of the general population do. Instead most – Wason (1968) suggested about 65% – incorrectly select the 3 card.

The selection task has spawned a phenomenal number of investigations in the psychological literature that have closely replicated Wason’s findings in these abstract, non-deontic settings, but which have given a wide range of different explanations for the results. They include confirmation bias, matching bias, Bayesian optimal data selection and relevance theory (see Sperber, Cara & Girotto (1995) for a review). It has even been suggested that the fundamental issue was the ‘defectively’ educated participants (Bringsjord, Noel & Bringsjord 1998).

This paper does not explore these theories in any depth. Instead we merely note that no explanation is generally accepted, and that the reasons behind Wason’s results remain unclear. However, given that no existing study has noted a significant difference in performance between those of differing subject backgrounds (and none has investigated mathematicians’ performance), we note that the main theories explain the *uniform* poor performance.

The goal of the current study was to compare the performance of mathematicians and non-mathematicians on the task. If the received view of logic’s place in mathematical thought is accurate, one might expect the mathematics undergraduates to perform significantly better than the general population, and the mathematicians to perform nearly flawlessly.

METHODOLOGY

In order to maximise the sample size available to us, we used an internet based survey. There were three categories of participant: mathematics undergraduates, mathematics academic staff and history undergraduates. All were from the University of Warwick. The historians were selected as representatives of the general population, as it was assumed their degrees would have little or no explicit teaching of logic. It is worth noting that this is in keeping with the practice of other researchers in the field: in many studies the population consists of psychology or other undergraduates. We are not aware of any studies conducted with more representative samples of the general population.

E-mails were sent to all members of the stated populations at Warwick, asking them to participate. If they agreed, they accessed a website which contained the following instructions:

> Four cards are placed on a table in front of you. Each card has a letter on one side and a number on the other.
>
> You can see:

> Here is a rule: *"every card that has a D on one side has a 3 on the other."*
>
> Your task is to select all those cards, but only those cards, which you would have to turn over in order to discover whether or not the rule has been violated.

This wording is identical to that used by Wason (1969). Once the subjects had submitted their answers, the webpage recorded five pieces of data: the subject's answer, whether or not they had seen the task before, which group the subject was from, the time, and their IP address. The answers from people who had seen the task before were deleted – very few (<2.5%) fell into this category. Participation rates were high: 260 maths students (34% of the whole population), 21 maths staff (24%) and 123 history students (23%) took part. These figures are particularly impressive when compared to the limited sample sizes available to Wason (1968, 1969) and other pre-web experimenters.

Using the internet to conduct research brings problems as well as benefits. The experimenter has severely limited control on the conditions the subject took part in. Whether they were in a quiet office or a busy internet café is uncertain. Perhaps the biggest problem with the method, however, is that of multiple submissions. There is no foolproof way of preventing subjects submitting their answers more than once. We followed Reips (2000) who suggested logging the IP address of the subject. This isn't a watertight method; often users have dynamic addresses – each time they go online they are assigned a different one. But by logging the time and the IP address of subjects, it was possible to catch those who resubmitted in quick succession (there seemed to be only one case of this, and his or her answers were deleted).

In the end the main defence against resubmissions is simply that subjects have no incentive to do so, it offers them nothing and, to avoid being caught by the IP address log, it is very costly in terms of time. Indeed, one experiment (in the days when dynamic IP addresses were rare) put the resubmission figure at 0.5% of total submissions (Reips 2000, p.105). In another study, Krantz & Dalal (2000) compared the results of twenty internet based surveys with their laboratory counterparts and

	Maths Students		Maths Staff		History Students	
D	92	35%	5	24%	27	22%
DK	1	0%	0	0%	0	0%
D3	15	6%	1	5%	41	33%
D7	76	29%	9	43%	10	8%
DK3	0	0%	1	5%	2	2%
DK7	34	13%	3	14%	1	1%
D37	8	3%	2	10%	8	7%
DK37	21	8%	0	0%	23	19%
non-D	13	5%	0	0%	11	9%
n	260		21		123	

Table 1: Answer Selections.

found a remarkable degree of congruence between the two methodologies. As a result of these factors, it seems clear that the benefits of using the web in this piece of research substantially outweigh the disadvantages.

RESULTS

The results of the study are shown in table 1.

The results for the historians are distributed in a similar fashion to those from Wason's (1968) original research on the 'general population'. This fact can only boost confidence in the methodology that we used.

It can clearly be seen that the maths students do indeed perform significantly differently to the history students (χ^2=95.9, p<0.001). However, although the mathematics students have a significantly higher success rate (χ^2=20.8, p<0.001), they still don't perform at all well. Less than a third of students – and less than half of staff – managed to identify the correct answer. Interestingly, a χ^2 test does not reveal a significant difference between the performance of the mathematicians and that of the mathematics students (χ^2=1.21, 5% level ≈ 3.8), though this may be due to the small numbers of maths staff taking part.

Looking carefully at the results reveals that not only did the mathematicians perform better than the non-mathematicians, but that they seemed to make *different* errors. This result is easy to see when the number picking each selection is expressed as a percentage of their group's incorrect answers only (see table 2). The history sample followed the pattern set by previous work: those that failed to choose the correct answer tended to pick D3, D or DK37. Those in the mathematics sample

	Maths Students	Maths Staff	History Students
D	50%	42%	24%
DK	1%	0%	0%
D3	8%	8%	36%
D7	-	-	-
DK3	0%	5%	2%
DK7	18%	25%	1%
D37	4%	17%	7%
DK37	11%	0%	20%
non-D	7%	0%	10%

Table 2: Incorrect Selections.

(both staff and students) who failed to find the correct answer were much more likely to select the D card on its own.

In the Wason Selection Task, the choice of each card corresponds to one of four logical inferences or common fallacies. Given the statement *every card that has a D on one side has a 3 on the other* (corresponding to $P \Rightarrow Q$), choosing the D card (corresponding to *P*) in the expectation of 3 (*Q*) on the other side suggests an appreciation of modus ponens. Choosing the K card (not-*P*) in the expectation of something other than a 3 (not-*Q*) suggests the fallacy of denying the antecedent. Choosing the 3 card (*Q*) in the expectation of a D (*P*) suggests the fallacy of affirming the consequent and choosing the 7 card (not-*Q*) in the expectation of something other than a D (not-*P*) suggests an appreciation of modus tollens. As well as comparing the frequency of each selection of cards, the results can be analysed in terms of the suggested inferential appreciations or fallacies (see table 3).

	Maths Students	Maths Staff	History Students	Inference or fallacy
D	95	100	91	modus ponens
K	25	19	26	denying the antecedent
3	20	19	62	affirming the consequent
7	57	67	40	modus tollens
n	260	21	123	

Table 3: The percentage of each group selecting each card.

The differences between the populations in their ability to recognise the relevance of the 3 card (and therefore their awareness of the logical fallacy of affirming the consequent) are stark. Nearly two thirds of historians selected it, whereas only a fifth of mathematicians thought it necessary. The other main difference between the groups was in recognising the validity of the modus tollens argument (this is the logical form of a contrapositive argument).

So, while it appears that mathematicians are significantly better at the selection task than non-mathematicians, from the point of view of their experience of learning and using logic, their performance is remarkably poor. Less than a third of maths students – and half of maths staff – answered correctly. These findings are somewhat surprising; no existing theory of performance on the selection task would seem to explain them. Our results thus raise two important questions:

- What are the features of mathematical cognition that allow mathematicians to perform significantly better than the general population? Why are they much less likely to make the standard mistake of selecting D and 3?

- What accounts for the unexpectedly poor performance of the mathematicians? If the role of logic in mathematics is as crucial as undergraduate textbooks would suggest, why didn't more respondents find the correct answer?

We suggest the role of error checking in mathematics may provide a potential answer to the first of these questions.

ERROR CHECKING IN MATHEMATICS

In his celebrated essay on mathematical invention Jacques Hadamard wrote:

> Good mathematicians, when they make [errors], which is not infrequent, soon perceive and correct them. As for me (and mine is the case of many mathematicians), I make many more of them than my students do; only I always correct them so no trace of them remains in the final result. (Hadamard 1945, p.49)

The importance of mathematical error checking was confirmed by Markowitz & Tweney (1981). In an empirical study of the behaviour of mathematicians when testing a conjecture, they found that 'disconfirmatory strategies' play a much greater role in mathematics than in the physical sciences. Thus we can claim that while mathematicians frequently make errors, in contrast to non-mathematicians they are highly skilled at detecting and correcting them. This error-correcting might provide a tentative explanation for our results.

We suggest that, along with the rest of the population, the initial reaction of the mathematicians to the Wason Selection Task would be to choose D and 3. If Hadamard is correct with his idea that mathematicians are significantly more adept at error checking, the typical mathematician would check their answer carefully, and quickly see that the 3 card was unnecessary. At this point they could do one of two things. Happy that they had corrected an error, they might stop checking and select just the D card (35% of students and 24% of staff selected the D card only). Or; checking their new answer carefully, they might realise that the 7 card was crucial and amend their answer accordingly. Perhaps after a further error check revealing no mistakes, D and 7 would be selected (29% of students and 43% of staff made this selection).

This chain of events would explain the two biggest selections by the mathematicians; that of D and D & 7. We conducted a brief pilot qualitative study involving clinically interviewing students as they attempted to solve the task. The small amount of data we have collected provides some support for our hypothesis: mathematics students initially choosing D and 3, pausing, rejecting the 3, pausing again and then choosing the 7. We are currently working on a larger scale qualitative study.

A reasonable way of testing the error checking hypothesis would be to time the responses of subjects. If it was true that an extra level of error checking caused the subjects to answer correctly, then one might expect them to take slightly longer. Such an experiment might prove to be a useful test of our tentative theory.

The second of our questions is related to the first. Faced with the initial selection of D and 3, there are two errors to be spotted: the incorrect selection of 3, and the failure to select 7. The data would seem to suggest that our sample was much better at finding the first of these errors than the second. The reasons for this are rather harder to pin down.

Amongst others, Johnson-Laird & Byrne (1991) suggest that human deduction fits remarkably poorly with formal logic. However, with the exception of Markowitz & Tweney's (1981) work, there has been little empirical research into exactly how professional mathematicians use fundamental logical ideas (such as disconfirmation) in mathematics. Previous work that describes the mathematical discovery process has mostly relied upon personal experience (Hadamard 1945, Tall 1980) and historical analysis (Lakatos 1976). Our data suggests that the role of logic in mathematicians' reasoning may be somewhat more subtle than previously thought.

When replying to our request for comments after the experiment one senior mathematics lecturer wrote:

> I don't think many of us think about the logical definition of $P \Rightarrow Q$ when writing out a proof in a research paper. The truth table for $P \Rightarrow Q$ is not very intuitive.

Could it be the case that there is a significant difference between the intuitive grasp of logic and the formal theory in the case of highly successful, professional mathematicians? If so, one has to ask why such emphasis is placed upon formal logic in first year undergraduate courses.

Acknowledgments:

We would like to thank Giles "the power" Taylor for his assistance with technical matters.

References:

Bringsjord, S., Noel, R. & Bringsjord E. (1998), In defence of logical minds *in* 'Proceedings of the 20th Annual Conference of the Cognitive Science Society', Hillsdale, New Jersey:Laurence Erlbaum, pp. 173-178.

Devlin, K. (2000), *The Maths Gene: Why everybody has it, but most people don't use it*, London: Weidenfeld & Nicholson.

Johnson-Laird, P. and Byrne, R. (1991) *Deduction*, London: Lawrence Erlbaum.

Hadamard, J. (1945), *The Psychology of Invention in the Mathematical Field*., Princeton University Press.

Hoyles, C. & Küchemann, D. (2002), 'Students' understanding of logical implication', *Educational Studies in Mathematics* **51**(3)*,* 193-223.

Krantz, J.H. & Dalal, R. (2000), "Validity of web-based psychological research", *in* M.H.Birnbaum (ed) *Psychological Experiments on the Internet*, San Diego: Academic Press, pp. 35-60.

Lakatos, I. (1976), *Proofs and Refutations: The logic of mathematical discovery*, Cambridge: CUP.

Markowitz, L. & Tweney, D. (1981), "An investigation of the behaviour of mathematicians engaged in testing a conjecture", presented at Midwestern Psychological Association, Detroit, MI; May 1981.

Oakhill, J., Johnson-Laird, P. & Garnham, A. (1989). "Believability and syllogistic reasoning" *Cognition* **31**, 117-140.

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. 35-60.

Sperber, D., Cara, F. & Girotto, V. (1995), 'Relevance theory explains the selection task', *Cognition* **57** 31-95.

Stewart, I. & Tall, D. (1977), *The Foundations of Mathematics*, Oxford: OUP.

Tall, D. (1980), 'The anatomy of a discovery in mathematics research', *For the Learning of Mathematics* **1**(2), 25-34.

Wason, P. (1968), 'Reasoning about a rule', *Quarterly Journal of Experimental Psychology* **20**, 273-281.

Wason, P. (1969), 'Regression in reasoning?', *British Journal of Psychology* **60**, 471-480.