CHARACTERISING MATHEMATICAL REASONING: STUDIES WITH THE WASON SELECTION TASK

Matthew Inglis & Adrian Simpson

University of Warwick

This paper analyses the nature of mathematical cognition with reference to the recently developed dual process theory account of reasoning. We briefly summarise dual process theory, and then present evidence from a study where mathematics students, mathematics staff and history students were asked to solve the Wason selection task, a standard logic question from the psychology literature. The mathematicians gave a dramatically different range of answers to the non-mathematicians. Using interview data from the same study, we suggest that one of the major differences between mathematical and non-mathematical thought is the ability, or willingness, to use System 2 processes whilst reasoning.

DUAL PROCESS THEORIES OF REASONING

Recently, psychologists have proposed that there are two distinct cognitive units that deal with reasoning. Roughly speaking the first corresponds with intuitive thought, and the second with abstract reasoning. Although there are many different versions of similar theories (e.g. Evans & Over, 1996; Skemp, 1979) the generic terminology System 1 and System 2, adopted by Stanovich & West (2000), has become commonplace.

System 1 is characterised by processes that are quick, operate in parallel and are highly context specific. These processes are almost entirely subconscious in nature, only the end product is deposited in the conscious brain. The system is independent of language, is old in evolutionary terms and is also present in animals. System 1 is believed to be a large collection of subsystems that operate autonomously. Some of these subsystems are believed to be innate, whilst others may have been acquired by an experiential learning mechanism (Stanovich, 2004).

System 2, on the other hand, is slow, operates in serial and allows for non-contextualised hypothetical reasoning. It is controllable and conscious, has evolved relatively recently and is unique to humans. It is this part of the brain that allows humans to construct complex abstract simulations that are context independent and depersonalised. Fluency with System 2 is often measured using reasoning tests, and tends to be correlated with measures of general intelligence (although it is perhaps not surprising that one form of reasoning test correlates with another). System 2 is also involved in expressing the output of System 1, and it has the ability to monitor and, possibly, override these intuitive responses, although, as we shall see, this does not always happen.

Although System 1 is innate, it can be developed over time through experience. For example, it has long been recognised that chess grandmasters, as well as having superior analytical skills, have a different way of ‘seeing’ the chess board to amateur players (e.g. de Groot, 1965). Their experience of chess playing has altered their System 1 heuristics as well as developed their System 2 analytical skills. (See Evans, 2003 for a full review of dual process theories).

EMPIRICAL EVIDENCE FOR THE DUAL PROCESS ACCOUNT

Since the sixties there has been mounting evidence that participants in reasoning experiments do not always respond in a normative manner. Take, for example, the ‘Linda’ problem (Tversky & Kahneman, 1983). In this task participants were told:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.

Participants were then given eight possible descriptions of her present employment and activities, and were asked to rank them in order of probability. Intriguingly, 85% ranked “Linda is a bank clerk and active in the feminist movement” as more probable than “Linda is a bank clerk”. Clearly, such a ranking is impossible. Tversky & Kahneman named this the ‘conjunction fallacy’, and explained it by noting that Linda resembles the prototypical feminist bank clerk more than she resembles the prototypical bank clerk.

Tversky & Kahneman, however, also noted that it would be unreasonable to claim that their (highly educated) participants had conceptions of probability that were largely based on resemblance to prototypical examples. Instead, the dual process account argues that the Linda task’s standard response comes from System 1. It is intuitive, automatic and more concerned with social data than formal logic. System 2 cues the opposite response, that which realises that $P(A)$ cannot possibly be less than $P(A \cap B)$. Individuals who respond with the conjunction fallacy, then, fail to successfully use System 2 to monitor and correct their intuitive System 1 output.

There is also neuropsychological evidence that supports the dual process account of reasoning. Goel & Dolan (2003) used fRMI brain scans whilst participants took standard reasoning tasks. They found that responses traditionally associated with System 1 were related to activity in the ventral medial prefrontal cortex, whereas the logically correct System 2 responses originated in the right inferior prefrontal cortex, an entirely different part of the brain. They concluded that System 1 reasoning was influenced by emotional processes.

THE SELECTION TASK

More important evidence that supports the dual process account of reasoning comes from the Wason selection task (Wason, 1968). First published in the sixties, the selection task has become the most investigated task in the whole psychological literature on reasoning.

Participants in the task are shown a selection of cards, each of which has a letter on one side and a number on the other. Four cards are then placed on a table:

The participants are 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 logically 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 make the ‘standard mistake’ of picking the D and 3 cards. Indeed, Wason (1968) suggested that about 65% incorrectly select the 3 card.

There is a vast psychological literature that has attempted to explain why so few participants make the correct selection. Forty years of research has failed to reach a consensus, and the detailed reasons behind Wason’s original results remain highly controversial. There are, however, some findings that have been found to be very stable, one of the most robust is the so-called ‘matching bias’ effect.

Evans & Lynch (1973) varied the structure of the task by rotating the presence of negatives in the rule (for example, they used rules such as “not D $\Rightarrow$ 3” and “D $\Rightarrow$ not 3” as well as the original “D $\Rightarrow$ 3”). They found that participants tended to select the cards that were mentioned in the rule, regardless of the presence of negatives. For the participants, the relevant cases seemed to be those that had the same lexical content as the propositional rule. Evans (2003) argues that this tendency, which has become known as ‘matching bias’, is a built-in heuristic in System 1. By the dual process explanation, the intuitive response, coming from System 1, is to select the D and 3 cards. It is argued that the standard mistake originates from participants using System 2 to merely rationalise and articulate this selection. As with the Linda problem, it is only if System 2 is actively and effectively monitoring System 1 that the logically correct answer (D and 7) can be produced. System 2 needs to *reason* rather than merely *rationalise* if the logically correct answer is to be found.

To reemphasise, the dual process account suggests that the standard mistake can be explained by a two part process: Firstly, card selections are determined entirely by System 1. Secondly, any System 2 processing that occurs is aimed at rationalising and articulating System 1’s output. There is empirical evidence to support this hypothesis. A key prediction of this account is that participants will spend more time inspecting the cards that they select than those that they reject (as System 2 will be rationalising the selections). This was experimentally verified by Evans (1996) using a computer based mouse hovering technique, and by Ball, Lucas, Miles & Gale (2003) who used a sophisticated eyeball tracking system to measure inspection times.

The two routes that lead to the correct answer and to the standard mistake are shown in figure 1. It should be noted that figure 1 is somewhat misleading as, as mentioned above, System 2 is used when rationalising and expressing any output from System 1. However, in figure 1 it is shown as playing no part in route 1 to emphasise that it is not involved in the *reasoning* process.

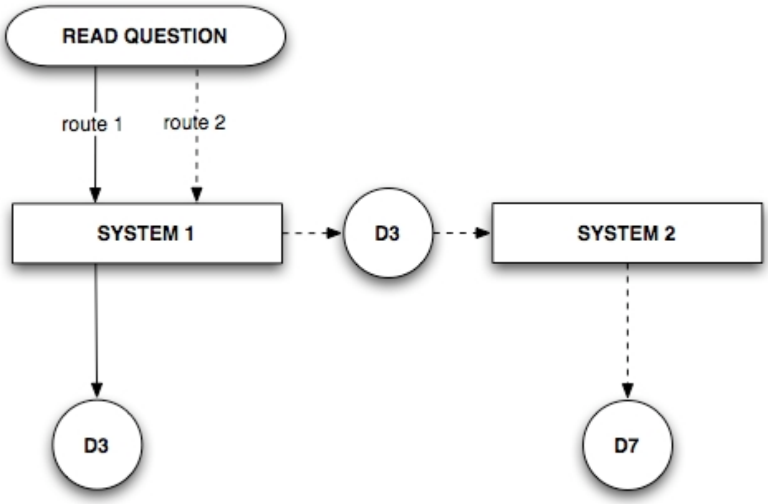

Figure 1: the routes that lead to the standard mistake and to the correct answer.

One of the most famous and striking results in the selection task literature is that performance can be dramatically facilitated by placing the task in a thematic context (e.g. Cosmides, 1989; Wason & Shapiro, 1971). Dual process theory explains this facilitation by suggesting that, in these thematic contexts, both System 1 and System 2 output the same answer. That is to say that only the abstract version of the task requires System 2 reasoning for its solution.

Its important to note that dual process theory is, to a large extent, neutral regarding the many competing theories that have attempted to explain performance on the Wason selection task. For example, mental models theory, mental logic theory, the pragmatic reasoning schemas theory and relevance theory can all be comfortably situated within a dual process framework.[1] Each of these theories can be seen as attempting to explain *why* either System 1 or System 2 produce the output that they do. In this sense, dual process theory is less a theoretical framework, and more a framework for theoretical frameworks.

As the above discussion illustrates, some of the most common reasoning misconceptions from the psychology literature can be explained by the failure of System 2 to adequately monitor and correct the intuitive output generated by System 1. This paper is concerned with the interplay between the System 1 and System 2 reasoning of successful mathematicians. Has day-to-day exposure of deductive

[1] It is less easy to situate the so-called social contract theory (e.g. Cosmides, 1989; Leron, 2004) within a dual process framework, as it would appear to dramatically underestimate the role of System 2. In any case, this explanation has been heavily criticised in recent years, and there is mounting empirical evidence that appears to contradict it (e.g. Sperber & Girotto, 2002).

reasoning modified mathematicians' System 1 heuristics in a manner similar to the chess masters, or is System 2 the key to the differences between mathematical and non-mathematical reasoning?

METHODOLOGY

We administered a version of the Wason selection task to three groups: mathematics undergraduates, mathematics academic staff and history undergraduates. The history students were used as a control group, as it was assumed their degree would contain little or no mathematical reasoning. It is worth noting that this is common practice, in many comparable studies the general population is represented by psychology undergraduates. We are not aware of any selection task studies that have used a more representative group.

We adopted an internet methodology. Potential participants were sent emails explaining the experiment and asking them to take part. If they agreed, they clicked through to a website which recorded their answer, whether or not they had seen the task, and their IP address. Further discussion of the methodology used, including its drawbacks, can be found in Inglis & Simpson (2004). However, our sample has since been expanded to include additional students from another high ranking UK university.

The precise wording we used was identical to Wason's (1969):

> 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.

Along with the quantitative based study we conducted a small number of clinical interviews with both mathematicians and historians who were not involved in the quantitative study. A standard 'think aloud' protocol was used.

RESULTS & DISCUSSION

The results are shown in table 1. Looking at the table reveals that there are significant differences between the mathematics and history students' range of answers (χ^2=100, df=8, p<0.001, ϕ=0.480). The mathematicians have a significantly higher success rate, although given the supposed importance of logic in mathematics, at only 28% for students and 43% for staff, it is perhaps surprisingly low.

Critically, for our analysis, very few of the mathematics students or staff made the 'standard mistake' – that of selecting the D and 3 cards. Only 6% and 5% of maths

	Maths Students		Maths Staff		History Students	
D	108	35%	5	24%	27	22%
DK	1	0%	0	0%	0	0%
D3	19	6%	1	5%	41	33%
D7	88	28%	9	43%	10	8%
DK3	0	0%	1	5%	2	2%
DK7	40	13%	3	14%	1	1%
D37	10	3%	2	10%	8	7%
DK37	23	7%	0	0%	23	19%
non-D	23	7%	0	0%	11	9%
n	312		21		123	

Table 1: The selections made by the different groups.

students and staff respectively made this mistake, compared to 33% of history students. Instead, by far the most common mistake made by mathematicians was to select only the D card.

Recalling that dual process theory claims that the ‘standard mistake’ originates in System 1, this result would appear to have deep implications about the nature of mathematical reasoning. What explains the absence of the ‘standard mistake’ from the mathematicians’ range of answers? What explains the proliferation of the D selection from mathematicians?

There would appear to be two reasonable hypotheses:

Hypothesis 1. Exposure to mathematics on a daily basis modifies System 1 heuristics in a manner similar to the chess grandmasters described earlier.

Hypothesis 2. Mathematicians’ System 1 tends to operate in the same way as for the general population. But exposure to mathematics on a daily basis results in an increased tendency to use System 2 for monitoring and possibly modifying the output of System 1.

The two hypotheses are illustrated in figure 2. As before, System 2 is not shown to be involved in hypothesis 1 to emphasise that it is not involved in the *reasoning* process.

If the first hypothesis were the case, it would appear that matching bias – the heuristic that seems to be responsible for the ‘standard mistake’ – is a casualty of exposure to mathematics. We are currently working on a study that is designed to test whether this mathematicians’ exhibit matching bias in more varied contexts.

If, however, the second hypothesis is the case, and given that a majority of both mathematics students and staff failed to select the correct answer, it would seem that the System 2 processes of many mathematicians are not as efficient at detecting logical mistakes as one might think.

The ‘think aloud’ protocol data that we collected provides some support for the second of these hypotheses. In the following extract Wolfgang, a postgraduate mathematical economics student, has just been handed a version of the selection task.

Wolfgang: [Reads the task] Hmmm. OK, first I know that I don't have to… that I definitely have to turn over the right one [the D].

Interviewer: The D?

Wolfgang: Yeah, the D, because that's the rule of course […] and the next thing, is the 3 has to be checked for, because when they are… [long pause] no, no, I don't have to, no I don't have to.

Interviewer: You don't have to? Why not?

Wolfgang: Because… because there can be another symbol, and if it does not violate the rule then there is a D and every other thing is not said, so it is not necessary to do it.

Wolfgang's initial reaction is that he needs to pick the D and the 3 cards. If dual process theory is to be believed, this is the result of the presence of a matching bias heuristic within Wolfgang's System 1. However, quickly and unprompted, he uses his System 2 to monitor and check his intuitive System 1 reaction. After several seconds of System 2 reasoning he corrects his answer, and decides that the 3 card is unnecessary, even if his explanation is somewhat incoherent. He eventually went on to find the correct answer.

Wolfgang was typical. Victor, a mathematics undergraduate at the end of his first year, responded similarly:

Victor: [Reads the question] Umm… you should turn over the D and the 3… then… they're the only ones.

Int.: Why?

Victor: Because it doesn't say anything about K's or 7's.

Interestingly, Victor even gives a matching bias based explanation for his selections. In fact every mathematician interviewed (n=6) gave D and

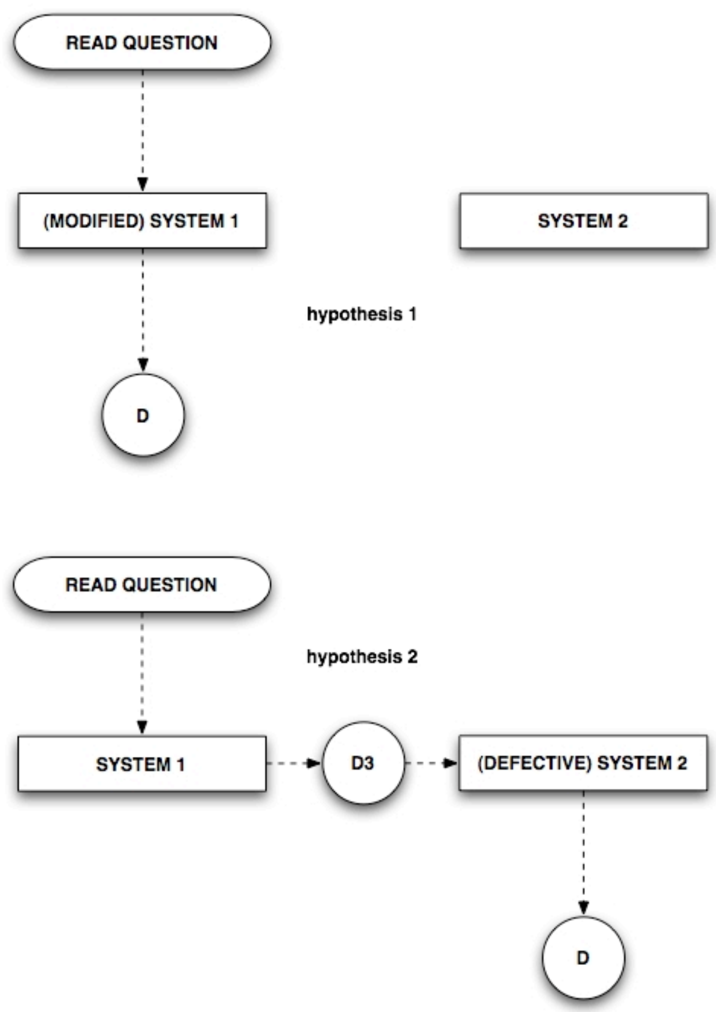

Figure 2: Two hypotheses that explain mathematicians' selection of the D card.

3 as their initial answer. However, they were able to use System 2 to modify and correct this mistake. Sometimes this took prompting. For example, here is Yasmin (a first year mathematics undergraduate) responding to the task.

Yasmin	I’d take the D and I’d have, umm, the 3. Because, if that had the D on the other side then the rule would be… [laughs].
Interviewer	What are you thinking?
Yasmin	I've got myself tied up in a knot!
Interviewer	OK, well, not to worry. […] What’s your thinking about the 3?
Yasmin	Umm, OK, I would leave the 3, you wouldn't need to turn that over. Because, yeah... but I would need to turn the other two over, I'd have to turn over the K and the 7 to see if there was a D on the other side.

Evidence from the, admittedly small scale, qualitative study would seem to indicate quite clearly that the initial reactions of mathematicians are the same as everyone else. Their System 1 heuristics are as influenced by matching bias as the general population. Given this, the quantitative data garnered from the internet study would suggest that mathematicians tend to be highly effective at using System 2 to monitor and criticise their System 1 reasoning.

However, given that the mathematicians appear to be using System 2, why do a large percentage of them fail to select the correct answer, D and 7? If they are using System 2 to detect and correct System 1’s mistaken decision to select the 3 card, why don’t they overwhelmingly detect the necessity of the 7?

Despite the proliferation of anecdotal or historical analyses of how mathematicians do mathematics (e.g. Hadamard, 1945; Lakatos, 1976; Tall, 1980), there is very little empirical research in the area. Based on the limited evidence presented here, it would seem that the understanding and use of the modus tollens deduction ($P \Rightarrow Q$, not Q, $\therefore$ not P) is not nearly as important to success in mathematics as one might think. More research is needed in this area.

CONCLUSIONS & SUMMARY

The quantitative data presented in this paper clearly indicates that there is a significant difference between mathematical and non-mathematical cognition. The mathematics students’ and staffs’ range of responses to the Wason selection task was significantly different to those from history students, who were taken to represent the general population. In particular both mathematics staff and students were significantly more likely to make the correct selection, and significantly less likely to make the standard mistake. Instead, the mathematicians’ standard mistake was to select the D card on its own.

We argue that this result has implications about the nature of mathematical cognition. In particular, if a dual process account of reasoning is adopted it would seem that either mathematicians have a modified System 1, or that they are much more likely to use System 2 to modify and correct their System 1 reasoning. Evidence from a small scale qualitative study provides some support for the second of these hypotheses. That is to say that one of the key differences between successful mathematicians and the general population is that the mathematicians appear more able to, or more willing to, activate their System 2 whilst reasoning.

However, the evidence also suggested that, for many mathematicians, their System 2 processes might not be as proficient at detecting some logical mistakes as one might think. Crucially though, since the mathematicians involved in this study were all highly successful, it would seem that this lack of proficiency does not have any serious effect upon their mathematical abilities.

Acknowledgement

We gratefully acknowledge the financial support of the Economic & Social Research Council.

References

Ball, L.J., Lucas, E.J., Miles, J.N.V. & Gale, A.G. (2003) "Inspection times and the selection task: What do eye movements reveal about relevance effects?", *Quarterly Journal of Experimental Psychology* **56A**(6), 1053-1077.

Cosmides, L. (1989) "The law of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task", *Cognition* **31**, 187–276.

de Groot, A. (1965) *Thought and choice in chess*, Mouton, The Hague.

Evans, J.St.B.T. (1996) "Deciding before you think: Relevance and reasoning in the selection task", *British Journal of Psychology* **87**, 223-240.

Evans, J.St.B.T. (2003) "In two minds: dual-process accounts of reasoning", *Trends in Cognitive Sciences* **7**(10), 454-459.

Evans, J.St.B.T. & Lynch, J.S. (1973) "Matching bias in the selection task", *British Journal of Psychology* **64**(3), 391-397.

Evans, J.St.B.T. & Over, D.E. (1996) *Rationality and reasoning*, Psychology Press, Hove.

Goel, V. & Dolan, R.J. (2003) "Explaining modulation of reasoning by belief", *Cognition* **87**, B11-B22.

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.

Inglis, M. & Simpson, A. (2004) "Mathematicians and the Selection Task", in M. Johnsen Hoines & A.B. Fuglestad (eds), *Proceedings of the 28th International Conference on the Psychology of Mathematics Education*, Bergen, Norway, vol. 3, pp. 89-96.

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

Leron, U. (2004) “Mathematical thinking and human nature: Consonance and conflict”, in M. Johnsen Hoines & A.B. Fuglestad (eds), *Proceedings of the 28th International Conference on the Psychology of Mathematics Education*, Bergen, Norway, vol. 3, pp. 217-224.

Skemp, R.R. (1979) *Intelligence, Learning and Action*, John Wiley & Sons, Chichester.

Sperber, D. & Girotto, V. (2002) “Use or misuse of the selection task? Rejoinder to Fiddink, Cosmides and Tooby”, *Cognition* **85**, 277-290.

Stanovich, K.E. (2004) *The Robot’s Rebellion: Finding Meaning in the age of Darwin*, Chicago University Press, Chicago.

Stanovich, K.E. & West, R.F. (2000) “Individual differences in reasoning: Implications for the rationality debate”, *Behavioural and Brain Sciences* **23**(5), 645-726.

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

Tversky, A. & Kahneman, D. (1983) “Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgement”, *Psychological Review* **90**, 293-315.

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

Wason, P.C. & Shapiro, D. (1971) “Natural and contrived experience in a reasoning problem”, *Quarterly Journal of Experimental Psychology* **23**, 63–71.