Increasing cognitive inhibition with a difficult prior task:

Implications for mathematical thinking.

Nina Attridge

University of Bath

Matthew Inglis

Loughborough University

Corresponding author: Nina Attridge, Centre for Pain Research, University of Bath, BA2 7AY, UK, n.f.attridge@bath.ac.uk.

Character count: 58,854

Abstract (142/150 words)

Dual-process theories posit two distinct types of cognitive processing: Type 1, which does not use working memory making it fast and automatic, and Type 2, which does use working memory making it slow and effortful. Mathematics often relies on the inhibition of pervasive Type 1 processing in order to apply new skills or knowledge that require Type 2 processing. In two studies we demonstrate that giving participants a difficult task (Raven's Matrices) before a task that requires the inhibition of intuitive responses (the Cognitive Reflection Test) significantly improves performance. Our findings suggest that encountering a difficult task that requires Type 2 processing before completing a task that requires inhibition of Type 1 processing may encourage an enduring 'Type 2' mind-set, whereby participants are more likely to spontaneously use Type 2 processing for a period of time. Implications for mathematics education are discussed.

Increasing cognitive inhibition with a difficult prior task: Implications for mathematical thinking.

1. Introduction

1.1. Dual-process theories

Dual-process theories of thinking have developed in many areas of cognitive psychology, including reasoning, decision-making and memory. Recently, dual-process theories have also been applied to mathematical thinking (e.g. Chernoff 2012, Gillard, Van Dooren, Schaeken and Verschaffel 2009a, Gómez-Chacon, García-Madruga, Vila, Elosúa and Rodríguez 2013, Leron 2010). Here, we argue that an intervention based on dual-process theory might improve students' mathematical thinking in many areas where intuition must be inhibited in favour of applying mathematical rules.

There are many versions of dual-process theory and while they vary in their specifics, there are some key shared characteristics. All propose at least two types of processing which we will refer to as Type 1 and Type 2 processing (Evans and Stanovich 2013). Type 1 processing does not require working memory, making it fast and automatic (e.g. recognising a friend's face). Type 2 processing depends on working memory, making it slower and more effortful (e.g. memorising a phone number). While Type 1 processing is constantly engaged and beyond our control, Type 2 processing can be engaged to differing levels depending on the individual and the situation. For example, individuals differ in their thinking disposition making them more or less willing to engage in effortful Type 2 thinking in general (Stanovich 2009). Individuals also differ in their capacity for Type 2 thinking, so that when it is engaged it may be more or less efficient between individuals (Stanovich 2009).

Situational factors such as time available (Evans, Handley and Bacon 2009) and perceptual difficulty of the task (e.g. hard-to-read fonts, Alter et al. 2007) can also influence the extent of Type 2 thinking. Bornemann, Foth, Horn, Ries, Warmuth, Wartenburger and van der Meer (2010) found that students with higher general cognitive abilities allocated more processing resources to a novel geometric task than their less able counterparts and showed better performance. However, they did not allocate more resources to a familiar algebraic-transformation task, yet still showed better performance. This suggests that high-ability individuals can engage different levels of Type 2 processing depending on the task at hand.

Type 1 thinking is thought to be evolutionarily old (Evans 2003), making it well equipped to deal effectively and efficiently with most tasks that we face. However, there are times when Type 1 thinking is not sufficient, or even counterproductive, and Type 2 thinking must be used in order to reach a goal, in particular, novel or modern tasks such as making an insurance claim, taking an exam, or proving a theorem. Of interest here are cases where Type 1 thinking and Type 2 thinking differ in their solutions to a problem, i.e. where intuition conflicts with reason. In these cases it is important for Type 1 intuitions to be inhibited in favour of Type 2 processing. There are three main perspectives on how this intervention occurs: the parallel-competitive, pre-emptive and default-interventionist perspectives.

Parallel-competitive theories (e.g. Sloman 1996) suggest that Type 1 and Type 2 thinking work on a problem simultaneously from the start. If they come to the same answer, the processing ends and the answer is given. If the processes give conflicting answers, Type 2 thinking can override Type 1 thinking, but because Type 1 thinking is much faster than Type 2 thinking it often wins the competition.

Pre-emptive theories suggest that it is decided at the outset of a task, based on its superficial characteristics, which type of processing will be used, and so there is never a conflict between different responses. For example, when faced with a syllogism with an unbelievable conclusion, it is decided on first reading that Type 2 processing should be used to evaluate it (Evans, Newstead and Byrne 1993).

Finally, default-interventionist theories (e.g. Evans 2006) suggest that Type 1 processing is the default way to solve any task, but in some cases Type 2 thinking can be engaged and potentially override the Type 1 response. In Evans's model, Type 1 thinking provides an answer which is subject to evaluation. This answer is often deemed to be plausible and is given as the response. Occasionally, however, the Type 1 answer is not found to be satisfactory and the problem is re-processed with Type 2 thinking.

In the current paper, we adopt a default-interventionist perspective on the interaction between Type 1 and Type 2 thinking. The nature of Type 2 thinking intervention from this perspective is discussed in more detail below in Section 1.3. First, however, we discuss the evidence that Type 1 and Type 2 processes can conflict in mathematics, and that incorrect answers from Type 1 thinking often win.

1.2 Cases of conflict in mathematical thinking

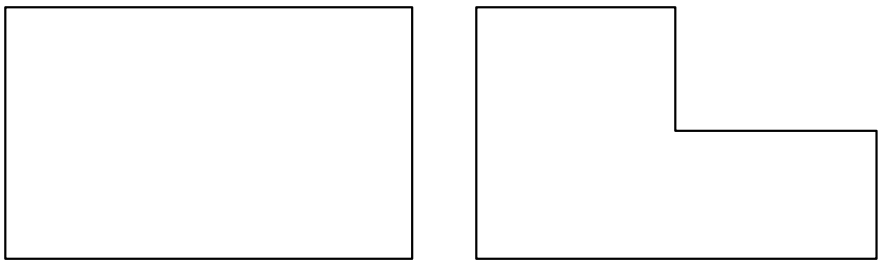
There are several areas of mathematics where Type 2 thinking is essential, but where intuition can get in the way. A first example is in dealing with rational numbers. Verschaffel, Van Dooren and colleagues (e.g. Obersteiner, Van Dooren, Van Hoof and Verschaffel 2013, Vamvakoussi, Van Dooren and Verschaffel 2012, Van Hoof, Lijnen, Verschaffel and Van Dooren 2013) have documented a 'Natural Number Bias', which occurs when knowledge of natural numbers interferes with

operations involving rational numbers (see also Christou 2015, this issue; Van Hoof, Janssen, Verschaffel and Van Dooren 2015, this issue). For example when comparing the fractions 2/5 and 2/7, students may have an intuition that the latter is larger than the former based on their long-held knowledge of natural numbers and the fact that as natural numbers 7 is larger than 5. This may not lead to an incorrect response, but the conflict between the intuitive response (Type 1 process) and knowledge of the correct response (Type 2 process) may be revealed in longer reaction times for correct answers to incongruent problems compared to correct answers to congruent problems (e.g. 2/5 vs 3/5, Van Hoof, Lijnen, Verschaffel and Van Dooren 2013). The natural number bias persists into adulthood, even in expert mathematicians (Obersteiner, Van Dooren, Van Hoof and Verschaffel 2013), and has been documented in fraction comparison and arithmetic operations (Vamvakoussi, Van Dooren and Verschaffel 2012).

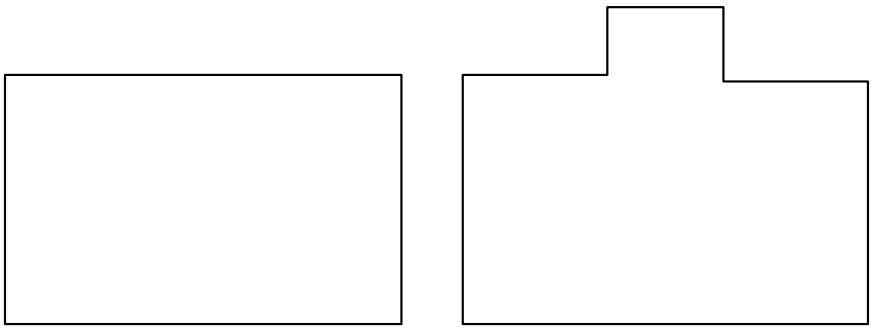
A second example where intuition can conflict with correct mathematical responses is in proportional reasoning. Gillard, Van Dooren, Schaeken and Verschaffel (2009b) investigated the tendency for adults to give proportional answers to nonproportional mathematics problems, such as “Ellen and Kim are running around a track. They run equally fast but Ellen started later. When Ellen has run 5 laps, Kim has run 15 laps. When Ellen has run 30 laps, how many has Kim run?”. The correct answer is 40 (e.g., “30 + 10 = 40” or “15 + 25 = 40”), but it is common for participants to give a proportional answer of “30 × 3 = 90” (Van Dooren, De Bock, Hessel, Janssens and Verschaffel 2005). Even trainee teachers were prone to this error (Cramer, Post and Currier 1993). Gillard et al argued that because proportional reasoning is such an important tool in everyday life and early education, students become very familiar and competent with it and become vulnerable to applying it

where it is inappropriate. By giving participants proportional and nonproportional problems either under a short time limit or under a working memory load, Gillard et al demonstrated that proportional errors stem from Type 1 thinking while overriding them depends on Type 2 thinking.

A third example of conflict between Type 1 and Type 2 processes occurs in geometry, where students tend to associate a larger shape area with a larger perimeter despite this not always being the case (see Figure 1). This is thought to be due to an intuitive *more A (e.g. area) – more B (e.g. perimeter)* rule (Babai, Shalev and Stavy 2015, this issue; Stavy and Tirosh 1996; Stavy, Tsamir and Tirosh 2002), which can lead to errors on incongruent problems such as those in Figure 1 A, or longer reaction



A) Incongruent perimeter comparison problem. The shape on the left has a larger area but equal perimeter to the shape on the right.



B) Congruent perimeter comparison problem. The shape on the right has a larger area and a larger perimeter compared to the shape on the left.

Figure 1. Example perimeter comparison problems where the area and perimeter are a) incongruent or b) congruent.

times if the rule is overridden in favour of Type 2 processing (Stavy and Babai 2010). Stavy and Babai demonstrated that, in educated adults, accuracy was lower for incongruent than for congruent problems, that correct responses to incongruent problems took longer than correct responses to congruent problems, and that correct responses to incongruent but not congruent problems activated brain areas associated with executive control (i.e. processes involved in managing cognition, such as inhibition). Furthermore, they demonstrated that an intervention designed to draw attention to the relevant variable (perimeter) significantly improved 14 and 15 year olds' performance on incongruent comparisons and increased their response times, suggesting an increase in Type 2 thinking. This finding is encouraging because it demonstrates that it is possible to increase reliance on Type 2 thinking in situations where Type 1 and Type 2 thinking conflict.

A fourth example of the conflict between intuition and reason in mathematics is in the area of probability. Gillard, Van Dooren, Schaeken and Verschaffel (2009c) tested a dual-process model of the bias to rely on frequency over probability when choosing between gambles. When participants are confronted with two possible gambles, such as those in Figure 2, and asked to choose the one with the best chance of picking a black marble, they may incorrectly choose the one with a larger number of black marbles over the one with a more favourable ratio of black marbles to white marbles. The finding that correct responses to incongruent items take significantly longer than intuitive responses suggests that the bias stems from Type 1 processing and that it can be inhibited and overridden by Type 2 processing, although this takes additional time.

A fifth and final example of the conflict between intuition and mathematics is the classic students and professors problem (Clement, Lockhead and Monk 1981), which has recently been interpreted in terms of dual process theory (Leron and Hazzan

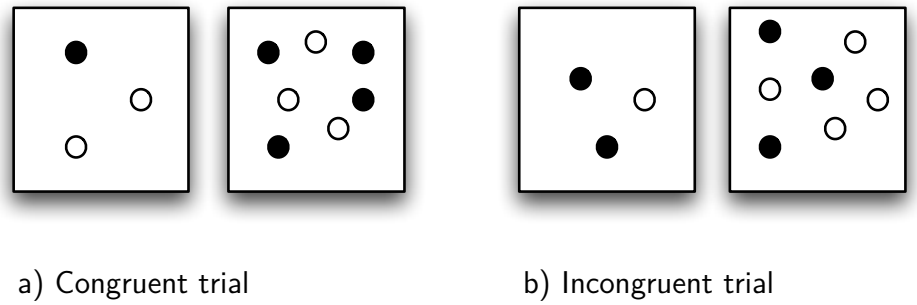


Figure 2. Example probability comparison problems where frequency and ratio of black marbles (targets) to white marbles (non-targets) are a) congruent or b) incongruent.

2006). The task is to write an equation for the statement "There are six times as many students as professors at this university" using *S* to represent the number of students and *P* to represent the number of professors. A common but incorrect response is "6*S* = *P*" (Clement et al. 1981). Leron and Hazzan suggested that Type 1 processing gives the pairing 6*S*, because of the phrase "six times as many students", and this is not overridden by Type 2 processing. Had the participants engaged Type 2 processing, they could have given the correct answer because they had the knowledge necessary to do so.

We have discussed several examples where intuition may conflict with reason in mathematical thinking (and other examples exist, see for example Lem 2015, this issue, on the law of large numbers). In each case, Type 1 processing leads to a response that is not always appropriate, and Type 2 thinking is necessary to override this response. A goal of mathematics education is to foster Type 2 thinking in cases such as these. However, the biases we have discussed are not short-lived problems that are overcome during formal education, rather they persist into adulthood reflecting the pervasiveness of Type 1 thinking and the difficulty of consistently

engaging Type 2 thinking. This leads us on to the question of how we decide when to use Type 2 thinking, and how we might help students to use more Type 2 thinking in cases where it is necessary.

1.3 Inhibiting Type 1 thinking errors

The recent adoption of dual process theories in mathematics education has led to a consideration of interventions to improve students' mathematical thinking. Leron and Hazzan (2006) have pointed out that training Type 2 thinking is not the solution, because in most cases where students show intuition-based errors they already have the knowledge and skills available to find the correct answer, they just fail to use them. Similarly, Type 1 thinking is difficult to manipulate (Lem, Onghena, Verschaffel and Van Dooren 2013, Tirosh, Stavy and Cohen 1998), as demonstrated by the fact that the biases discussed above persist throughout schooling and can be observed in educated adults.

Instead of targeting interventions at Type 1 or Type 2 processes themselves, the solution may lie in encouraging students to question their Type 1 thinking and more often override it with Type 2 thinking. There is evidence that individuals may be aware when there is a conflict between their intuition and reason, which provides an opportunity for us to encourage them to act on this conflict more often (De Neys, Vartanian and Goel 2008; Lubin, Simon, Houdé and De Neys 2015, this issue). This is the perspective we take on inhibition in the current paper, i.e. we are interested in increasing the tendency to inhibit the output of Type 1 processing.

Tirosh et al (1998) observed students' tendency to follow two intuitive rules: 'everything comes to an end' and 'everything can be divided'. They then attempted to reduce students' use of these rules by presenting them with conflicting arguments

from two hypothetical students and asking them to determine which, if any, of the arguments were correct and to justify their answer. This could be expected to draw the students' attention to the conflict and encourage them to evaluate the two arguments with Type 2 thinking, but unfortunately, the intervention did not significantly change the students' reliance on the intuitive rules at post-test.

On the other hand, Houdé et al. (2000) successfully trained students to avoid matching bias[1] by warning them about the bias and instructing them in how to avoid it. The training and testing utilised different tasks, so this was not a simple case of showing participants how to complete the task correctly. Rather, it was designed to make participants aware of the trap they may fall into in order to help them avoid it. As well as a large shift from incorrect to correct responding, the intervention resulted in a shift in the brain areas that were active during reasoning.

Similarly, Lem et al (2013) attempted to reduce undergraduate students' intuitive errors in their interpretation of box plots by giving them a very detailed and explicit explanation of how to correctly interpret a box plot. Although the students did give more correct responses after the intervention than before, their reaction times for correct responses to counterintuitive problems suggested that they still encountered incorrect heuristics (as mentioned above, Type 1 processing is very difficult to manipulate); they were simply able to override them.

Although these interventions had some success, a downside is that they are task-specific. The same issue applies to Stavy and Babai's (2010) intervention that was discussed above, and ideally an intervention to encourage the use of Type 2 thinking would be task-general so that it could be applied across several areas of mathematics, and even in non-mathematical domains.

Recently, several authors have proposed metacognitive processes that monitor the output of Type 1 processing and determine when Type 2 thinking will be engaged (e.g. Thompson 2009, Stanovich 2009), and these processes may provide a domain-general target for interventions. Thompson proposed that Type 1 outputs come with an associated 'feeling of rightness' (FOR). The FOR is an affective construct with little cognitive content. Thompson suggested that if an answer that came from Type 1 thinking has a high FOR, the individual will be likely to accept that answer without questioning it. If the answer comes with a low FOR, however, the individual will be more likely to inhibit it and tackle the problem with Type 2 thinking. This suggests that individuals have a threshold for what they consider to be a sufficient FOR, and this may vary between and within individuals.

Thompson, Prowse Turner and Pennycook (2011) found evidence that FOR affected behaviour in four reasoning tasks. Participants were asked to provide a quick initial response to each question (heavily influenced by Type 1 processes), indicate how 'right' the answer felt (FOR), and were then allowed as much time as they wanted to come to a final answer (potentially, although not necessarily, based on Type 2 thinking). Initial answers that were associated with a low FOR were associated with longer re-thinking times for the second answer, suggesting that low FOR prompts more Type 2 thinking. Low FOR was also associated with a higher probability of changing one's answer when allowed time to re-think, indicating that FOR influenced trust in the intuitive answer at a behavioural level.

Stanovich (2009) similarly proposed a reflective level of cognition, which is superordinate to what he calls the autonomic (Type 1) and algorithmic (Type 2) levels (metacognition is generally considered to be a form of Type 2 thinking and Stanovich does not dispute this, but he describes the algorithmic and reflective aspects of Type 2

thinking as subordinate and superordinate, respectively). The reflective level consists of thinking disposition, which is an individual's tendency to value effortful thought and their willingness to engage it, and this can also vary between individuals. Those who value effortful thought would be more likely to inhibit the output of their Type 1 processing when working in a domain that they consider important, opting instead to evaluate the problem with Type 2 thinking.

Thompson's FOR threshold and Stanovich's reflective level (which are not mutually exclusive, but may be considered parts of the same metacognition construct) provide potential targets for interventions to increase the inhibition of Type 1 thinking and the adoption of Type 2 thinking. As mentioned above, situational factors can determine whether Type 1 or Type 2 thinking is used, including time available, instructions and perceptual difficulty of the task, and these factors may be working through such mechanisms. For example, Alter et al (2007) showed that presenting tasks in a difficult-to-read font increased Type 2 thinking, and they suggested that this was because the perceptual difficulty of the task prompted participants think that the task itself was going to be difficult. This may have had the effect of altering their threshold for FOR or priming their reflective level of cognition to override 'easy' answers. One of the tasks that Alter at al used was the Cognitive Reflection Test (CRT, Frederick 2005, Figure 3), which is useful for investigating Type 2 process engagement because the task strongly prompts intuitive but incorrect answers, and the correct answers are fairly simple to calculate if Type 2 thinking is engaged and the intuitive answer inhibited. Participants' responses are therefore thought to be a good reflection of whether or not they have inhibited their Type 1 intuitions.

Thompson et al (2013) distinguished between two types of fluency that affect participants' level of Type 2 thinking on tasks such as the CRT where there is a

1. A bat and a ball cost £1.10 in total. The bat cost £1 more than the ball. How much does the ball cost? _____ pence
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes
3. In a lake there is a patch of lily pads. Every day the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days

Figure 3. Frederick's (2005) Cognitive Reflection Test. The correct answers are 5 pence, 5 minutes and 47 days, but participants often give the intuitive answers 10 pence, 100 minutes and 24 days.

conflict between Type 1 and 2 processing: answer fluency, i.e. the ease with which an answer comes to mind (which is directly related to FOR), and perceptual fluency, i.e. the ease with which the task as a whole is perceived on a superficial level (varying, for example, due to font manipulations). The authors investigated the effects of both types of fluency on the rate of Type 2 processing across seven experiments, and consistently failed to replicate Alter et al's (2007) findings that perceptual fluency impacted on extent of Type 2 processing. Answer fluency, on the other hand, consistently affected processing. This suggests that perceptual fluency may not be a reliable method of encouraging students to engage Type 2 processing, but since answer fluency is a property of the processes that lead to an answer, it is not likely to be malleable by short-term interventions.

1.4. The present studies

Here we aimed to test a new method of manipulating students' inclination to inhibit Type 1 thinking on the CRT; a method that does not rely on changing the perceptual or answer fluency of the problem. The method was intended to influence the individual's metacognitive processes, such as the reflective level of cognition or the threshold for FOR (i.e. the level of 'rightness' that was required of an answer before the individual would be willing to accept it without further scrutiny). We gave participants an explicitly difficult task, Raven's Advanced Progressive Matrices (RAPM), before giving them the CRT. RAPM items do not prompt a strong intuitive answer so Type 2 processing is essential (unless an answer is picked at random), and because the RAPM items are difficult the level of Type 2 processing required is high. We hypothesised that this would have a similar but more reliable effect on responses than perceptual fluency manipulations because participants experience actual task-processing difficulty immediately prior to completing the relevant task, as opposed to perceptual difficulty at a purely superficial level. By targeting the metacognitive processes responsible for intervention in this way, we aimed to put participants in an enduring mind-set of needing to engage Type 2 thinking, in the process inhibiting Type 1 processing.

If successful, this intervention could help students to overcome incorrect intuitions in mathematical problems such as those described above. Because Type 1 thinking is so pervasive, students would still be aware of the intuitive answers to problems, but they may experience a conflict between this answer and their metacognitive processes which encourage them to think the problem through more carefully. Not only would they be more likely to give correct answers, but the experience of the conflict and of the benefit of engaging more effortful thinking (i.e.

coming up with a different answer to their first intuition) may teach them to question their intuitions more frequently in the future.

In Study 1 we asked participants to complete the CRT either before or after completing a subset of 18 items from RAPM with a 10 minute time limit, and we show that completing the RAPM items first significantly improved performance on the CRT. In Study 2 we asked participants to complete the CRT after completing zero, one, three or five RAPM items, and show that one RAPM item was sufficient to improve CRT performance compared to zero, while additional items did not provide any additional benefit (i.e. the effect was not cumulative).

2. Study 1

2.1 Method

2.1.1 Design

The study followed a between-subjects experimental design. The independent variable was order of tasks (RAPM first or CRT first), and the dependent variable was number of correct answers on the CRT task. The study was approved by the Loughborough University Ethics Approvals (Human Participants) Sub-Committee.

2.1.2 Participants

Participants were 61 second year undergraduate students from engineering courses at Loughborough University. Participants took part during a lab session for an introductory statistics module and later analysed the data for their coursework assignment. All participants gave informed consent for their data to be used for research purposes.

2.1.3 Tasks

Participants completed an 18-item subset of Raven's Advanced Progressive Matrices (RAPM, Raven, Raven and Court 1998) with a 10 minute time limit. This subset was selected for an undergraduate population in previous research by removing items on which there tended to be floor or ceiling effects (Sá, West and Stanovich 1999). In each item a 3×3 matrix displayed a pattern with a piece missing. Participants decided which of eight spare pieces completed the pattern, and pressed the corresponding number key on a standard keyboard. Each item was presented until response. If participants finished all 18 items in less than 10 minutes the task ended, otherwise the task was automatically terminated after 10 minutes.

Participants also completed the three-item Cognitive Reflection Test (CRT, Frederick 2005). Each item was presented individually on screen until response, in a random order. Participants typed in their answer and pressed the spacebar to continue. Both tasks were administered using E-Prime 2.0 (Schneider, Eschman and Zuccolotto 2002).

2.1.4 Procedure

The study took place in groups in a computer lab with participants attending one of three sessions. Approximately 20 participants attended each session. Participants were seated at computers and instructed to work alone and in silence. The experiment program randomly allocated each participant to one of the two order conditions; 32 participants did the CRT first and 29 did the RAPM first. After the CRT and RAPM tasks, participants completed several other tasks that are not relevant to the current research question. Because the RAPM and CRT tasks were completed first, performance could not be influenced by the subsequent tasks.

2.2 Results

The mean number of CRT items answered correctly by each group is shown in Figure 4. The number of correct answers to the CRT was compared across groups using the non-parametric Mann-Whitney U test, due to the limited scale (0-3). The RAPM first group scored significantly higher (M = 1.72, SD = 1.10) than the CRT first group (M = 1.16, SD = 1.08), $U(61) = 597.0$, $z = 1.99$, $p = .047$, indicating that completing the RAPM task first significantly improved CRT performance, as predicted[2]. The effect size was medium, Cohen's $d = .51$, with CRT scores being 48% higher after 10 minutes of RAPM items compared to the CRT being completed first. Conversely, RAPM scores did not significantly differ between the RAPM first (M =

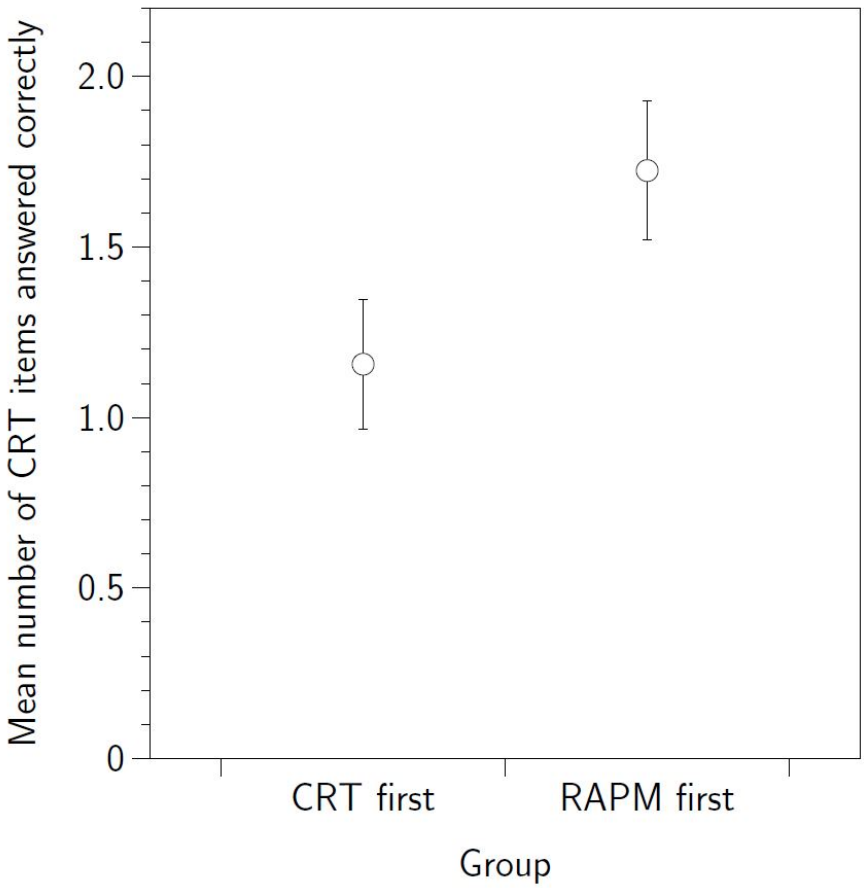


Figure 4. Mean number of correct responses to the three CRT questions in the RAPM first and CRT first groups. Error bars represent ±1 standard error of the mean.

9.10, *SD* = 2.79) and CRT first groups (*M* = 8.47, *SD* = 3.41), $t(59) = .790$, $p = .432$.

There was a significant correlation between RAPM and CRT scores, $r_s(61) = .385$, p = .002, which replicates previous findings that performance on the CRT is related to individual differences in intelligence – more intelligent individuals are more likely to engage in effortful thinking, or to do so more effectively, than less intelligent individuals. However, the similarity of RAPM scores in the two groups suggests that this effect did not contribute to the difference in CRT scores between groups.

2.3 Discussion

Participants completed a difficult subset of RAPM items before or after completing the CRT. The CRT questions prompt intuitive but incorrect responses which can be overridden by engaging effortful thinking to evaluate one's answers. Those who completed the RAPM items first scored significantly higher on the CRT, i.e. they were more likely to engage effortful thinking to overcome their intuitions. The effect did not occur in the opposite direction – RAPM scores did not differ depending on whether or not the CRT was completed first.

RAPM items cannot be solved with Type 1 thinking alone and as such require Type 2 thinking. It appears that this experience has a lasting effect on the likelihood of Type 2 thinking intervention on a subsequent task for which Type 2 thinking is not necessary in order to produce a response (in so far as Type 1 thinking does provide a response, albeit an incorrect one). Our findings suggest that it is possible to manipulate the likelihood of engaging Type 2 thinking, but our intervention was fairly intensive (up to 10 minutes long). In Study 2, we investigated whether the same effect could be found with a more practical intervention, and whether the intensity of the intervention had a cumulative effect on subsequent task performance.

3. Study 2

We found encouraging results using a 10 minute difficult task, but 10 minutes could be considered a long time in the context of a school lesson or university lecture. Next, we investigated whether giving participants only 1, 3 or 5 RAPM items would have a similar effect. We did not include an 18 item condition for two reasons: first, the study was conducted online and we had no direct contact with participants, meaning so many items may lead to a disproportionately high dropout rate compared to the other conditions, and second, because so many items would be impractical for the classroom and the aim of the study was to determine whether the intervention was effective at more practical intensities.

In Study 2 we were also interested in whether the effect of prompting Type 2 thinking would be cumulative. In other words, would asking participants to tackle multiple RAPM items be more effective at prompting Type 2 thinking than asking them to tackle one?

3.1 Method

3.1.1 Design

The study was administered online and followed a between-subjects experimental design. The independent variable was number of RAPM items with four conditions: zero, one, three or five. The dependent variable was number of correct answers to the CRT. The study was approved by the Loughborough University Ethics Approvals (Human Participants) Sub-Committee.

3.1.2 Participants

Participants were 489 adults (307 male, 179 female, 3 unspecified), aged 20 – 65 (*M* = 28.36, *SD* = 9.09), who were recruited through Amazon Mechanical Turk and completed the study online in exchange for 20 cents (USD).

3.1.3 Amazon Mechanical Turk

Amazon Mechanical Turk (www.MTurk.com) is an open online marketplace for recruiting 'workers'. It allows 'requesters' to post tasks that can be completed at a computer (such as surveys, writing, transcription, data entry and experiments), which registered workers can be paid to complete. Over 100,000 workers are registered from over 100 counties, and between them they complete thousands of tasks every day (Pontin 2007). Workers can browse and select tasks that they want to complete, and after satisfactory completion they are paid. The rate of pay is usually low (often around $1 per hour), but research suggests that workers are intrinsically motivated (Buhrmester, Kwang and Gosling 2011). Buhrmester et al have shown that MTurk can be used to obtain high-quality research data quickly and inexpensively. Participants recruited via MTurk are more demographically diverse than those recruited at universities (Buhrmester et al. 2011, Paolacci, Chandler and Ipeirotis. 2010), the rate of pay does not affect data quality (Mason and Watts 2009), and the data are as reliable as those collected through traditional methods (Buhrmester et al. 2011, Paolacci et a., 2010).

Web-based research methods more generally have been found to provide results consistent with those found using traditional methods (Gosling, Vazire, Srivastava and John 2004; Krantz and Dalal 2000), and there are many advantages which make this a useful and promising tool for recruiting research participants (Reips 2000). However, there are some limitations to internet studies, such as higher dropout rates (Birnbaum 2004, Piper 1998), lack of control over the testing

environment (Hecht, Oesker, Kaiser, Civelek and Stecker 1999), and increased opportunity for dishonesty (Suri, Goldstein and Mason 2011).

3.1.4 Tasks

Five of the 18 RAPM items used in Study 1 were selected on the basis of having similar levels of difficulty (items 15, 16, 17, 19 and 20 from Raven's Advanced Progressive Matrices Set II), which was determined by scores in a previous data set (Attridge and Inglis 2013). Participants saw zero, one, three or five of these items and, as before, selected one of eight pieces to complete the pattern. The three CRT items were presented one at a time, in a set order, and participants typed in their response. Both tasks were programmed in Wextor (Reips and Neuhaus, 2002).

3.1.5 Procedure

Participants were directed to the study site via Amazon Mechanical Turk. On the first webpage they saw study information and checked a box to indicate either that they wanted to seriously participate or that they wanted to just browse the webpages. Next, they were asked to report their age and sex. Following this, they were randomly allocated to one of the four RAPM conditions. The tasks were then presented, and finally, a debrief page gave a summary of the study's purpose and provided the contact details for the researchers.

3.2 Results

These were 115 participants in the zero RAPM group, 141 in the one RAPM group, 124 in the three RAPM group, and 109 in the five RAPM group. The mean number of correct answers given by participants in each group is shown in Figure 5. Due to the non-parametric nature of the CRT data, we compared the median number of correct answers to the CRT across the four conditions (zero, one, three or five

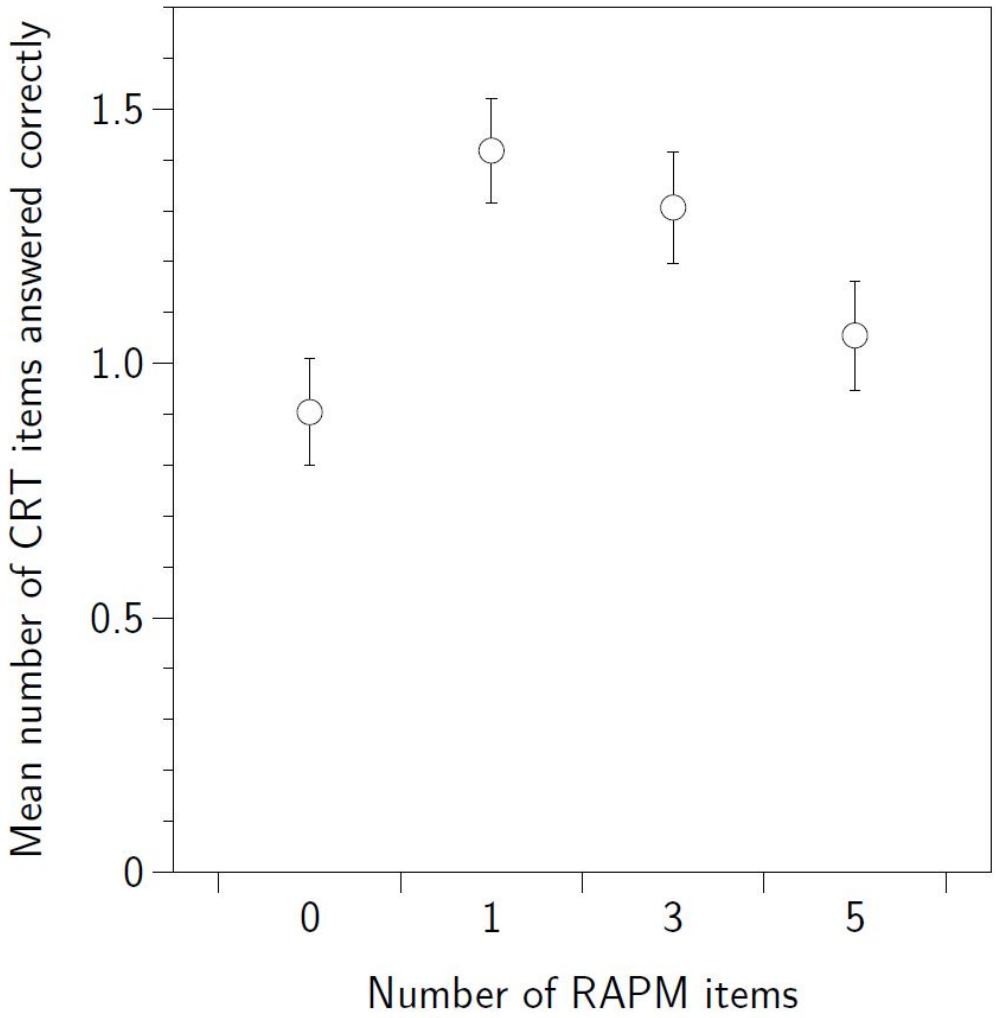


Figure 5. Number of correct responses to the three CRT questions in the 0, 1, 3 and 5 RAPM items first conditions. Error bars represent ±1 standard error of the mean.

items) with a Kruskal-Wallis test. This revealed a significant main effect of Condition, $H(3) = 14.00$, $p = .003$[3], which was further investigated with a series of Mann-Whitney U tests to compare each pair of conditions. We used Bonferroni correction to control for multiple comparisons, which made the critical alpha 0.008 (i.e. 0.05/6). Participants who saw one RAPM (median = 2, mean = 1.42, *SD* = 1.23) scored significantly higher (58%) than those who saw none (median = 0, mean = 0.90, *SD* = 1.12), $U(256) = 6216.5$, $z = 3.39$, $p = .001$. Those who saw three (median = 1,

mean = 1.31, *SD* = 1.22) scored marginally higher (46%) than those who saw none, $U(239) = 5811$, $z = 2.63$, $p = .009$, but they did not score higher than those who saw one, $U(265) = 8318$, $z = .71$, $p = .477$. Participants who saw five RAPM items (median = 1, mean = 1.06, *SD* = 1.12) did not score higher than those who saw none, $U(224) = 5737.5$, $z = 1.17$, $p = .240$, or three items, $U(233) = 6020.5$, $z = 1.51$, $p = .132$ and they scored non-significantly lower (25%) than participants who saw one RAPM item, $U(250) = 6444$, $z = 2.29$, $p = .022$.

3.3 Discussion

A large sample of adults completed the CRT online after zero, one, three or five RAPM items. Those who saw one RAPM item scored significantly higher than those who saw none, and there was no additional benefit to seeing three or five RAPM items. This indicates that the effect of RAPM items prompting Type 2 engagement is not cumulative. In fact, there was a non-significant trend for more items to result in lower CRT scores. One possible explanation for this might be the nature of the data collection. Participants were taking part online for very little reward, and they potentially lost interest when they were asked to tackle five difficult RAPM items, leading to the slight drop off in CRT performance.

These results are very promising – they suggest that seeing just one difficult problem is enough to have an enduring effect on the likelihood of Type 2 thinking intervention. Such a technique could potentially be used in mathematics classrooms to encourage students to use Type 2 thinking when tackling problems that are known to elicit incorrect intuitive responses. Our intervention does not change the Type 1 processing, nor does it provide additional knowledge or strategies to improve the quality of Type 2 thinking (as discussed above, the quality of Type 2 thinking is not

usually the issue, rather the unwillingness to engage it). Instead, our intervention appears to have made our participants more likely to question the output of their Type 1 thinking and to re-process the problems using Type 2 thinking. This means that the intervention is domain-general. Some potential criticisms are addressed below.

4. General discussion

There are many areas of mathematics where incorrect intuitions can interfere with the application of available knowledge and procedures. Here, we investigated a method of encouraging the inhibition of intuition in favour of more effortful processing from a dual-processes perspective. The dual-processes perspective posits that automatic Type 1 processing provides solutions to all kinds of tasks we face on a day-to-day basis. In some instances, this processing is not sufficient to reach a correct solution and effortful Type 2 processing must be used. In two studies we investigated the effectiveness of asking participants to complete an overtly difficult task (requiring Type 2 processing) before completing a task that strongly prompts incorrect intuitive responses. In Study 1, we observed that participants who completed 10 minutes of Raven's Advanced Progressive Matrices (RAPM) before completing the Cognitive Reflection Test (CRT) scored significantly higher on the CRT than those who completed RAPM second. In Study 2, it was found that even one RAPM item significantly improved performance on the CRT compared to no RAPM items.

This intervention was based on recent advancements of dual-process theories that have described metacognitive mechanisms which monitor the output of Type 1 processing and prompt their inhibition when it is deemed necessary (Stanovich 2009, Thompson 2009). We suggest that the intervention examined here may have been effective via such a mechanism. RAPM items do not prompt an intuitive response –

they are abstract, visuo-spatial and difficult. Type 2 processing is essential in order to come to an answer, even an incorrect answer, and this experience may have primed participants' metacognitive control processes to deem that Type 2 processing was necessary on the following task as well. This potential mechanism for our findings should be examined in future research by attempting to directly measure the metacognitive control processes, for example, using the methods of Thompson, Prowse Turner and Pennycook (2011).

There are several encouraging features of this intervention: it is domain-general (i.e. not tied to a specific mathematical bias), it is quick (even one RAPM item was effective), and the effect is fairly large (in Study 2, the mean score after one RAPM item increased by nearly 58% compared to zero RAPM items and the median score increased from 0 to 2 on a scale of 0-3). Previous interventions to encourage the inhibition of Type 1 biases on mathematical tasks have been either not effective or effective but task-specific. For example, Tirosh et al's (1998) intervention for the intuitive 'everything comes to an end' and 'everything can be divided' rules, which was not effective, and Houdé et al.'s (2000) matching bias intervention, which was effective but task-specific.

The intervention used here targets the mechanism that decides when a Type 1 intuition will be inhibited and Type 2 processing engaged. This processing architecture is thought to be general to all of cognition, meaning that an effective intervention of this type should be useful in all areas of mathematics (and even more broadly) where there is a conflict between intuition and reason. For example, when students are introduced to fractions or decimals, and there is a risk of their natural number knowledge interfering, it may be useful for them to complete a RAPM item at the beginning of each lesson or task (provided that they already have sufficient

knowledge of fractions/decimals to know that they behave differently to natural numbers).

Outside of mathematics, an intervention such as the one presented here could potentially help people make better decisions in their day-to-day lives, for example, when buying cigarettes, alcohol or junk food, gambling, investing money in the stock market, or choosing between insurance policies. In each case decisions can be influenced by non-relevant factors through Type 1 thinking, such as logos (Janiszewski and Mayvis 2001), brand advertising (Friese, Wänke and Plessner 2006), or the fluency of stock names (Alter and Oppenheimer 2006, Oppenheimer 2008). Completing a RAPM item first may help consumers to override the influence of these factors through Type 1 thinking, and avoid making a bad decision. For example, gamblers could be required to correctly solve a RAPM item before placing a bet with an online bookmaker, and the extra Type 2 thinking may lead to a more thorough consideration of the likelihood of losing money, and subsequently to the decision to not place a bet. Cigarette packets could feature an explicitly difficult RAPM item, similar to current interventions that require health warnings to be featured on cigarette packets (Hammond 2011), and this may encourage additional Type 2 thinking which may counteract the effects of branding on the decision to smoke. Of course, such implementations are highly speculative but there is the potential for interventions that prompt Type 2 thinking to have wide ranging impact.

The size of the effect found here was notable. In Study 1 CRT scores increased by 48% after 10 minutes of RAPM items, and in Study 2 they increased by 58% after only one RAPM item. However, there was non-significant drop-off in CRT scores after 5 RAPM items in Study 2. One possible explanation for this is that the study was conducted online and participants did not have any external incentive to

stay focused. It may be that after five difficult RAPM items participants' attention began to wander and their performance on the CRT suffered. Of course, this did not seem to occur in Study 1 where participants completed up to 18 RAPM items in up to 10 minutes. However Study 1 was conducted in a group setting in a computer lab as part of a course module, and the presence of other students, the researchers, and the formal setting may have facilitated participants' engagement with such a long and difficult task. It would be worthwhile to replicate Study 2 in a classroom setting to see whether the same drop-off trend is replicated, and if so whether it is significant. If our explanation is correct, then we would expect performance to be equal for the 1, 3 and 5 item conditions.

We have argued that the domain-generality of our intervention is a strength, but it could be argued that it is instead a weakness. One could argue that mathematics education should explicitly teach students about the tricky nature of tasks that tend to elicit Type 1 thinking errors, for example by warning them about common incorrect responses to particular tasks, and how each task should be solved correctly. Although our intervention is implicit (and could be used to complement rather than replace more explicit techniques), it is not necessarily experienced implicitly by participants. Our intervention made participants more likely to use Type 2 thinking on the CRT, but there is no reason to suppose that the intuitive responses did not still come to mind before being rejected and replaced with answers from Type 2 thinking, giving participants an awareness of the conflict between intuitive and effortful thinking. We suggest that our intervention made participants more aware of the insufficiency of their intuitive responses and the necessity of engaging Type 2 thinking. This personal experience of the issue may even be more effective at teaching students the tricky

nature of the task than simply telling them, although this is of course an empirical question.

Despite the advantages of the intervention tested here, namely the speed, effect size, and domain-generality, there are many remaining questions to be answered before this intervention could be deemed useful for the classroom. Firstly, we do not know the duration of the effect. Our results show that the need for Type 2 processing on the RAPM task carried over onto a subsequent task, but the CRT is a fairly short task and we do not know whether the same effects would be seen on a longer task, or after a delay. If the effect were found to be very time-limited, then it would not be suitable as an educational intervention.

Future research should also investigate the effect of a RAPM pre-task on some of the biases discussed in the introduction. The CRT has some similarities to these biases – it prompts intuitive but incorrect responses, the majority of participants have the necessary knowledge or ability to find the correct answer, and it is mathematical in nature – but it is only one task and the utility of the intervention should be established for a variety of mathematical tasks where Type 1 and Type 2 processes conflict. Different tasks elicit different heuristics, and some may be easier to inhibit than others. It is currently unclear whether our intervention would help performance on tasks with stronger heuristic responses than the CRT. Similarly, it would be interesting to compare the effects of other difficult tasks to the RAPM to establish the most effective tool.

Somewhat related to this point, it would be important for research to examine the intervention in different age groups, and it may well be the case that different intervention tasks would be suitable for different age groups. The RAPM task was effective in an adult sample, but it is designed for high-ability adults and may well be

too difficult to be effective in school students. It seems likely that a too-difficult task could cause students to give up rather than think harder. Raven's Coloured Progressive Matrices are designed for 5- to 11-year-old children, and Raven's Standard Progressive Matrices are designed for average-ability adolescents and adults, and within each set of Matrices there is a wide range of difficulty (hence the term 'progressive'), and so these tasks would be good starting points for establishing the most effective interventions in different age groups.

Finally, it would also be interesting to investigate individual differences in the efficacy of the intervention. It may be the case that students with higher cognitive ability see better effects because they have more resources available for Type 2 processing (Stanovich 2009) or are more flexible in their level of Type 2 engagement (Bornermann et al. 2010).

To conclude, we have introduced a potentially fruitful intervention based on dual-process theories that may encourage students to inhibit their intuitions in favour of using effortful thinking based on mathematical knowledge and rules.

Footnotes

1. Matching bias is a perceptual bias to focus on items that have been explicitly named. For example, in the Wason Selection Task, the rule 'if there is an A on one side of the card then there is a 3 on the other side' can prompt participants into focusing only on the 'A' and '3' cards and not considering the others (Wason 1968).
2. Condition had a marginally significant effect on the number of intuitive errors, $U(61) = 337.0$, $z = -1.90$, $p = .058$, and no effect on the number of 'other' (not intuitive but not correct) errors, $U(61) = 430$, $z = -.74$, $p = .462$.
3. Condition also had a significant effect on number of intuitive answers, $H(3) = 13.46$, $p = .004$, but not on the number of 'other' (not intuitive but not correct) answers, $H(3) = 2.31$, $p = .512$.

References

Alter, A. L., & Oppenheimer, D. M. (2006). Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences*, *103*(24), 9369-9372.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569.

Babai, R., Shalev, E., & Stavy, B. (2015). How a warning intervention aimed at activating inhibitory control mechanisms affects students' ability to overcome intuitive interference. *ZDM - The International Journal on Mathematics Education*, this issue.

Birnbaum, M. H. (2004). Human research and data collection via the Internet.*Annu. Rev. Psychol.*, *55*, 803-832.

Bornemann, B., Foth, M., Horn, J., Ries, J., Warmuth, E., Wartenburger, I., & van der Meer, E. (2010). Mathematical cognition: individual differences in resource allocation. *ZDM – The International Journal on Mathematics Education*, *42*(6), 555-567.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3-5.

Chernoff, E. J. (2012). Recognizing revisitation of the representativeness heuristic: an analysis of answer key attributes. *ZDM – The International Journal on Mathematics Education*, *44*(7), 941-952.

Christou, K. P. (2015). Natural number bias in operations with missing numbers. *ZDM - The International Journal on Mathematics Education*, this issue.

Clement, J., Lockhead, J., & Monk, G. (1981). Translation difficulties in learning mathematics. *American Mathematics Monthly, 88,* 286-290.

Cramer, K., Post, T., & Currier, S. (1993). Learning and teaching ratio and proportion: Research implications. *Research ideas for the classroom: Middle grades mathematics*, 159-178.

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

Evans, J. S. B., Handley, S. J., & Bacon, A. M. (2009). Reasoning Under Time Pressure. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *56*(2), 77-83.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241.

Evans, J. S. B., Newstead, S. E., & Byrne, R. M. (Eds.). (1993). *Human reasoning: The psychology of deduction*. Psychology Press.

Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. Psychonomic Bulletin & Review, 13 , 378-395.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 25-42.

Friese, M., Wänke, M., & Plessner, H. (2006). Implicit consumer preferences and their influence on product choice. *Psychology & Marketing*, *23*(9), 727-740.

Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009a). Dual processes in the psychology of mathematics education and cognitive psychology. *Human Development*, *52*(2), 95-108.

Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009b). Proportional reasoning as a heuristic-based process: time constraint and dual task considerations. *Experimental Psychology*, *56*(2), 92.

Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009c). Processing time evidence for a default-interventionist model of probability judgments. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1792-1797). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gómez-Chacón, I. M., García-Madruga, J. A., Vila, J. Ó., Elosúa, M. R., & Rodríguez, R. (2014). The dual processes hypothesis in mathematics performance: Beliefs, cognitive reflection, working memory and reasoning. *Learning and Individual Differences*, *29*, 67-73.

Hammond, D. (2011). Health warning messages on tobacco products: a review. *Tobacco Control*, *20*(5), 327-337.

Hecht, H., Oesker, M., Kaiser, A., Civelek, H., & Stecker, T. (1999). A perception experiment with time-critical graphics animation on the World-Wide Web. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 439-445.

Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., & Tzourio-Mazoyer, N. (2000). Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience*, *12*(5), 721-728.

Janiszewski, C., & Meyvis, T. (2001). Effects of brand logo complexity, repetition, and spacing on processing fluency and judgment. *Journal of Consumer Research*, *28*(1), 18-32.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35–60). San Diego, CA: Academic Press.

Lem, S. (2015). The intuitiveness of the law of large numbers. *ZDM – The International Journal on Mathematics Education*, this issue.

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). The heuristic interpretation of box plots. *Learning and Instruction*, *26*, 22-35.

Leron, U. (2010). The power of natural thinking: Applications of cognitive psychology to mathematics education. In *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 6, pp. 9-20). Columbus: Ohio State University.

Leron, U., & Hazzan, O. (2006). The rationality debate: Application of cognitive psychology to mathematics education. *Educational Studies in Mathematics*, *62*(2), 105-126.

Lubin, A., Simon, G., Houdé, O., & De Neys, W. (2015). Inhibition, conflict detection and number conservation. *ZDM – The International Journal on Mathematics Education*, this issue.

Obersteiner, A., Van Dooren, W., Van Hoof, J., & Verschaffel, L. (2013). The natural number bias and magnitude representation in fraction comparison by expert mathematicians. *Learning and Instruction*, *28*, 64-72.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237-241.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411-419.

Piper, A. I. (1998). Conducting social science laboratory experiments on the World Wide Web. *Library & Information Science Research*, *20*, 5-21.

Pontin, J. (2007, March 25). *Artificial intelligence: With help from the humans.* The New York Times. Retrieved from http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html. Accessed 10th April 2014.

Raven, J., Raven, J. C. & Court, J. H. (1998). Manual for raven's advanced progressive matrices and vocabulary scales. San Antonio: Pearson.

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-117). San Diego, CA: Academic Press.

Reips, U. D., & Neuhaus, C. (2002). WEXTOR: A Web-based tool for generating and visualizing experimental designs and procedures. *Behavior Research Methods, Instruments, & Computers*, *34*(2), 234-240.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 2.0). Computer software and manual. *Pittsburgh, PA: Psychology Software Tools Inc*.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? *In two minds: Dual processes and beyond*, 55-88.

Stavy, R., & Babai, R. (2010). Overcoming intuitive interference in mathematics: insights from behavioral, brain imaging and intervention studies. *ZDM – The International Journal on Mathematics Education*, *42*(6), 621-633.

Suri, S., Goldstein, D. G., & Mason, W. A. (2011, August). Honesty in an Online Labor Market. In *Human Computation: Papers from the 2011 AAAI Workshop (WS-11-11).*

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *In two minds: dual processes and beyond.* Oxford University Press, Oxford.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive psychology*, *63*(3), 107-140.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237-251.

Tirosh, D., Stavy, R., & Cohen, S. (1998). Cognitive conflict and intuitive rules, *International Journal of Science Education, 20*(10), 1257-1269.

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior*, *31*(3), 344-355.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction*, *23*(1), 57-86.

Van Hoof, J., Janssen, R., Verschaffel, L., & Van Dooren, W. (2015). Inhibiting natural knowledge in fourth graders: towards a comprehensive test instrument. *ZDM – The International Journal on Mathematics Education*, this issue.

Van Hoof, J., Lijnen, T., Verschaffel, L., & Van Dooren, W. (2013). Are secondary school students still hampered by the natural number bias? A reaction time

study on fraction comparison tasks. *Research in Mathematics Education*, *15*(2), 154-164.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, *20*(3), 273-281.