Thinking & Reasoning

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ptar20

Domain-specific experience and dual-process thinking

Zoë A. Purcell, Colin A. Wastell & Naomi Sweller

To cite this article: Zoë A. Purcell, Colin A. Wastell & Naomi Sweller (2021) Domain-specific experience and dual-process thinking, Thinking & Reasoning, 27:2, 239-267, DOI: 10.1080/13546783.2020.1793813

To link to this article: https://doi.org/10.1080/13546783.2020.1793813

THINKING & REASONING
2021, VOL. 27, NO. 2, 239–267
https://doi.org/10.1080/13546783.2020.1793813

Domain-specific experience and dual-process thinking

Zoë A. Purcell, Colin A. Wastell and Naomi Sweller

Department of Psychology, Macquarie University, Sydney, Australia

ABSTRACT
Prominent dual process models assert that reasoning processes can transition from effortful (Type 2) to intuitive (Type 1) with increases in domain-specific experience. In two studies we directly examine this automation hypothesis. We examine the nature of the relationship between mathematical experience and performance on the cognitive reflection test (CRT; Frederick, 2005). We test performance and response time at different levels of experience and cognitive constraint. Participants are required to complete a secondary task of varying complexity while solving the CRT. In Study 1, we demonstrate changes in thinking Type across real-world differences in mathematical experience. In Study 2, convergent with Study 1, we demonstrate changes in thinking Type across a mathematical training paradigm. Our findings suggest that for some individuals low experience is associated with Type 1 processing, intermediate experience is associated with Type 2 processing, and high experience is associated with Type 1 processing. Whereas, for other individuals low experience is associated with ineffective Type 2 processing, intermediate experience is associated with effective Type 2 processing, and high experience is associated with Type 1 processing.

ARTICLE HISTORY Received 13 May 2019; Accepted 6 July 2020

KEYWORDS CRT; expertise; training; reasoning; dual process theory

A novel problem or task may seem difficult at first, but with enough practice, it can become easy and routine. Practice and the process of learning is often accompanied by some mild cognitive unease and effortful thinking, but—over time—can eventuate in a transition from effortful to effortless thinking. Reasoning and thinking scholars, particularly dual process theorists, are interested in the differences between effortful (Type 2) and intuitive (Type 1) thinking, and have suggested that some Type 1 processes may be the product of Type 2 processes having been practiced to the point of

CONTACT Zoë A. Purcell zoe.purcell@mq.edu.au
Supplemental data for this article is available online at https://doi.org/10.1080/13546783.2020.1793813.

automation (e.g., Epstein, 1994; Evans & Stanovich, 2013; Kahneman, 2011; Sloman, 1996; Stanovich, 2018). This transition phenomenon (from Type 2 to Type 1), also known as the process of automation, is more thoroughly developed in theories such as Klein's (2003) naturalistic decision-making theory, Gigerenzer's (2007) fast and frugal heuristics theory, Reyna's (2012) fuzzy trace theory, and Wastell's (2014) emergent modularity theory. Recently, however, the issue of automation has come to the forefront of dual process theorising (De Neys & Pennycook, 2019; Stanovich, 2018).

The key tenet of dual process theory is that reasoning is achieved via two distinct types of processes: Type 1 which is automatic and intuitive, and Type 2 which is deliberate and effortful (e.g., Epstein, 1994; Evans & Stanovich, 2013; Kahneman, 2011; Sloman, 1996). Evans and Stanovich (2013) defined Type 1 processes as autonomous: they do not require controlled attention, are not dependent on input from high-level control systems, and—importantly—do not require working memory. Type 2 processes are characterised by the engagement of a general-purpose system, are responsible for cognitive decoupling and hypothetical thinking, and—in contrast to Type 1 processes—require working memory. Working memory is thought to be a relatively stable "hardware" of an individual's higher cognition, used to hold information that can be accessed and manipulated for a short time, but which is vulnerable to interference from competing cognitive tasks (e.g., Baddeley, 1986; Engle et al., 1999; Hambrick & Engle, 2002). The assertions that 1) reasoning becomes less dependent on working memory with practice, and 2) the distinction between thinking Types by the involvement (or not) of working memory are widely accepted within dual process theorising. However, until recently the integration of these assertions has been limited (for a review, see De Neys, 2017).

Recently, Stanovich (2018) discussed the interaction between thinking Type and experience. Building on classic models of automation (e.g., LaBerge & Samuels, 1974; Shiffrin & Schneider, 1977), he theorised a "mindware continuum" along which the reasoner's dependence on working memory shifts as domain-specific experience increases. The continuum depicts a person's reasoning as they move from low to high experience, that is, as they develop more advanced mindware. At novice stages of learning and experience, Stanovich suggests that the relevant mindware has not yet been developed. Further along the continuum, as the mindware is practiced and developed in long-term memory, the relevant procedures can be retrieved via Type 2 processes. Stanovich calls this intermediate phase the "Zone of Conflict" reflecting the role of cognitive unease postulated by hybrid dual process models that suggest that cognitive uncertainty may play a role in engaging Type 2 processes (see De Neys, 2012, 2014; Thompson, 2009; Thompson & Johnson, 2014). Once the reasoner

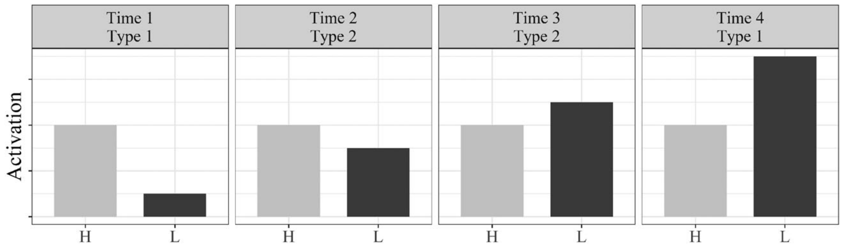


Figure 1. Conjectural model of the relationship between increases in experience, the relative activations of heuristic and logical processes, and the Type of reasoning that is engaged. H = Heuristic process, L = Logical process.

progresses through the intermediate phase or "Zone of Conflict", and the mindware becomes overly practiced, the relevant processes begin to display Type 1 characteristics. That is, the solution procedure for the task or problem can be retrieved and instantiated automatically and autonomously (i.e. without working memory). Stanovich effectively integrated the traditional models of automation with the contemporary dual process theories of reasoning. As such, he provided a solid theoretical grounding for the relationship between experience and Type of reasoning from a broad dual process perspective.

Although the proposed relationship between experience and thinking Type is consistent with a broad dual process approach to reasoning, it can be incorporated by some modern dual process models more easily than others. As Stanovich (2018) highlights, the models that are particularly suitable suggest that multiple Type 1 processes can be triggered at the outset of problem solving (De Neys, 2012; Handley & Trippas, 2015; Pennycook et al., 2015; Trippas & Handley, 2017). Of these, the logical intuition model is perhaps the most appropriate for the task as it contains a detailed framework for the mechanisms underlying the automation continuum. The logical intuition model asserts that, when faced with a problem, multiple Type 1 reasoning processes may be activated (De Neys, 2012, 2014). Although only the process with the highest activation is actioned or "wins out", the relative strength of each of the activated processes has cognitive ramifications. If two or more processes have similar levels of activation, conflict may occur which can manifest as a sense of cognitive unease or uncertainty (Bago & De Neys, 2017; De Neys, 2012, 2014). The more similar the relative activations, the more conflict a reasoner is expected to experience. This conflict, according to the model, may be involved in the engagement of Type 2—working memory dependent—processes. When considered alongside Stanovich's (2018) "mindware continuum," hypotheses can be formed about the interaction between experience, thinking Type, and task performance. This proposed phenomenon is reflected in Figure 1.

Figure 1 presents the hypothesised trajectory of reasoning Type as a person's experience increases. In line with the assertions from Stanovich's (2018) mindware continuum and De Neys' (2012, 2014) logical intuition model, the figure shows the activation of two processes over time, one leading to a heuristic response and the other leading to a logical response. The more training the logical process undergoes, the greater its activation potential and the more likely it is to "win-out" in the future. At Time 1, prior to training, the reasoner gives the heuristic response, experiences little conflict, and subsequently engages very little working memory—characteristic of Type 1 processing. After some training, at Times 2 and 3, the activation of the logical process comes closer to that of the heuristic response. At this intermediate stage, the similar activations of the two processes may lead the reasoner to experience more conflict and greater working memory engagement—characteristic of Type 2 processing and Stanovich's (2018) "Zone of Conflict". At Times 3 and 4, the reasoner gives the correct, logical response. At Time 4, the difference between activations is larger than at Time 3. Therefore, the reasoner experiences less conflict and working memory engagement decreases—characteristic, once more, of Type 1 processing. Although this example is highly simplified, it demonstrates the principle of transitioning from heuristic to logical processes via training.

There are three key phases of the mindware trajectory in regard to the expected shifts thinking Type across experience. First, the reasoner employs Type 1 processing; second, they engage Type 2 processing; and third, they employ Type 1 processing once again. The proposal that logical processes can become Type 1 is supported by the relationship observed between cognitive capacity and logical intuition (Raoelison et al., 2020; Thompson et al., 2018). Thompson et al. (2018) found that when high-capacity reasoners made belief-based judgements they were influenced by logical principles, while the opposite was true for low capacity reasoners. Relatedly, Raoelison et al. (2020) observed a stronger relationship (r=.42) between cognitive capacity and intuitive correct responses than capacity and deliberative correct responses (r=.13). These studies suggest that for high-capacity reasoners, some of whom may also have been high-experience reasoners, the logical responses were more accessible than the heuristic response. This is in line with the later stages of the proposed mindware trajectory (Stanovich, 2018).

Examining the proposed shifts between reasoning Type requires determining the Type of thinking in which a reasoner is engaged and whether that Type of reasoning shifts systematically over changes in experience. In the empirical components of this article, we achieve this by testing performance across different levels of cognitive constraint and experience. Previous studies have employed cognitive constraint techniques to determine a reasoner's thinking Type. These techniques are based on the dual process assertions that Type 2 processes demand more time and cognitive resources than Type 1 processes (Kahneman & Frederick, 2005; Kahneman, 2011). Hence, constraints that deprive the reasoner of time or working memory resources, like a time-limit or secondary task, have been used to determine whether the participant had required Type 2 thinking to solve a particular problem (e.g., Bago & De Neys, 2017, 2019; Thompson et al., 2011). These paradigms have been used to examine the nature of reasoning on the CRT (e.g., Bago & De Neys, 2017, 2019).

The CRT is a three-item test that commonly elicits heuristic, erroneous responses (Frederick, 2005). Consider the first CRT problem: "A bat and a ball together cost $1.10. The bat costs $1 more than the ball. How much does the ball cost?" This problem often prompts the heuristic response—10c, despite most responders having the ability to reach the logical solution—5c. As the name suggests, the CRT was originally proposed as a test of reflective thinking. Like many bias tasks, correct responses were interpreted as the result of Type 2 processes intervening and correcting the default Type 1 solution (e.g., Kahneman, 2011; Kahneman & Frederick, 2005; Toplak et al., 2011). However, studies that have employed cognitive constraints to examine the Type of thinking used to complete the CRT have yielded varied results (e.g., Bago & De Neys, 2017, 2019).

These studies have found evidence for logical responding on the CRT via both Type 1 and Type 2 processes. For example, Johnson et al. (2014, 2016) demonstrated that, on average, participants' performance on the CRT fell when they were required to complete a simultaneous visuospatial task. This indicates that at least some of the participants' correct responses were derived from Type 2 processes. Using an alternative cognitive constraint method, however, Bago and De Neys (2019) found that most participants giving correct responses on the bat-and-ball problem (the first item in the CRT) were doing so via Type 1 processes. Bago and De Neys employed a two-response paradigm in which participants were given two attempts to complete the bat-and-ball problem; the first attempt under cognitive load and time-pressure, and the second, without cognitive constraints. They found that most respondents who gave a correct answer at the second attempt, also gave the correct answer for their first attempt, indicating that these responses were the product of Type 1 processing. However, there was also a portion of participants who were only able to provide a correct response at the second attempt without cognitive restrictions; indicating that they had reached the solution via Type 2 processing. Together, these studies suggest that both Type 1 and Type 2 processes can be used to reach the logical solutions on the CRT.

Bago and De Neys (2019) considered that the correct responses provided under Type 1 conditions may be due to those reasoners having automated the required process. They acknowledged that automation is the goal of many learning contexts and that years of exposure during high school or other mathematical training may have helped these reasoners to do just that. In line with this interpretation and the theoretical assertions that thinking Type shifts across changes in domain-specific experience, we note that those participants who required Type 2 processes for success on the CRT may have been at an intermediate point of the automation such that they still needed to go through a process of deliberation that required cognitive resources. From the perspective of Stanovich's (2018) mindware continuum, these participants may have been in the "Zone of Conflict" in which the relevant procedures could only be retrieved via Type 2 processes. In the current article, we empirically explore the possibility that individual differences in mathematical experience may lead to systematic variability in their engagement of working memory when completing the CRT.

In two studies, we establish a trajectory of thinking Types across changes in experience. This is the first instance that we know of where experience manipulations have been used in conjunction with the CRT and cognitive constraints. We chose to focus on the CRT for several reasons. First, it is a mathematical problem-solving task, shown to correlate with numeracy (e.g., Cokely & Kelley, 2009; Liberali et al., 2012) and, therefore, lends itself to straightforward manipulations of mathematical experience. Second, most people give the incorrect responses on the CRT, leaving a greater capacity for improved performance via training (Frederick, 2005). Third, the CRT is a centrepiece of prominent reasoning models and features heavily in dual process literature. Therefore, it is an appropriate place to start when first testing the interaction between thinking Type and experience from a dual process perspective.

In both studies we examined the relationship between thinking Type and experience. Study 1 examined the relationship between real-world differences in mathematical experience, a manipulation of cognitive constraint, and performance on the CRT. Study 2 expanded on Study 1's design. It included a within-subject manipulation of experience (via training), a manipulation of cognitive constraint, a measure of working memory capacity (WMC; included as a continuous covariate), and measures of response time and performance on the CRT. Cognitive constraint was manipulated in both studies through the use of matrix memory tasks (Bago & De Neys, 2019). Study 1 employed a 2 (constraint) x 3 (experience) between-subjects design and Study 2 employed a 3 (constraint) x 3 (test point) x WMC (continuous covariate) mixed design. The hypotheses for each study are outlined in the corresponding sections below.

Study 1

In Study 1, experience was operationalised by classifying participants into one of three groups according to their university course or occupation (experience: low, intermediate or high). Cognitive constraint was manipulated by randomly allocating participants to one of two constraint conditions (constraint: load or no-load). We expected that, for CRT performance:

1. Participants with more mathematical experience would outperform those with less experience;
2. Participants in the no-load condition would outperform those in the load condition;
3. Participants with intermediate experience would be affected by the load to a greater extent than those with low or high experience.

Method

Participants and design

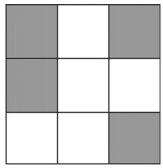
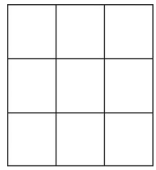
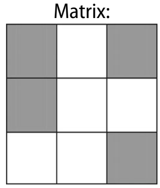
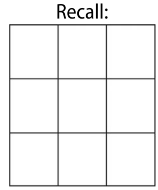
Study 1 employed a 2 (constraint) x 3 (experience) between-subjects design. Only one female participant qualified for the high experience group, therefore, this participant was excluded, and we used an all-male sample for the low and intermediate experience groups[1]. Final participants were 65 males, with ages ranging from 18 to 72 years (M_{age}=25.46, $SD = 12.96$). Participants were randomly allocated to the no-load ($N = 34$) or load ($N = 31$) condition. Low experience participants were 26 undergraduate psychology students at Macquarie University, Sydney (M_{age}=21.96, $SD = 10.65$). Intermediate experience participants were 24 undergraduate actuarial, science, or engineering students at Macquarie University, Sydney (M_{age}=20.33, $SD = 4.00$). High experience participants were 15 postgraduate mathematical students at Macquarie University, Sydney, or professional mathematicians (M_{age}=39.73, $SD = 15.94$). Undergraduate participants were recruited from Macquarie University, and non-students were recruited via a mathematics website. Psychology students were awarded course credit for participation. Non-psychology students were offered the chance to enter a draw for one of three $AU50 vouchers.

Materials

CRT. The three-item CRT was used (Frederick, 2005). There was no time limit imposed and a free-response format was used (i.e. not multiple

[1]Gender has been shown to affect CRT performance through mathematical anxiety (e.g., Frederick, 2005; see Morsanyi et al., 2018; Primi et al., 2018).

Table 1. Presentation of materials for (a) no load and (b) load conditions. Each of the four elements were presented on separate pages. The order of the elements differed between conditions.

	First	Second	Third	Fourth
(a) No load Condition	Information component (e.g., A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball.)	Matrix: 	Recall: 	Full item (e.g., A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?____)
(b) Load Condition	Information component (e.g., A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball.)	Matrix: 	Full item (e.g., A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?____)	Recall: 

choice). Participants entered a number in the units specified on screen. Total scores ranged from 0–3.

Constraint. Matrices were used as cognitive constraints. Participants were required to memorise 3 × 3 grids with four coloured squares presented for 800 ms (see Table 1). This task was adapted from previous dot matrix memory tasks for use on the Qualtrics survey platform. The matrices had four coloured squares that formed "three-piece" patterns (see Table 1; Bethell-Fox & Shepard, 1988; Verschueren et al., 2004). Matrices have been used effectively to impose constraints on reasoning processes (e.g., Bago & De Neys, 2019; Johnson et al., 2014, 2016). Matrix performance scores were calculated by scoring each coloured (or not) square as correct or incorrect. Scores are reported as percentages.

Numeracy. Hegarty et al.'s (1995) 12-item numeracy test was used to substantiate the categorisation of participants by course and occupation as a reflection of mathematical experience. There were no time limits imposed and a free-response format was used. An example of a PST item is: "At McDonald's, workers earn $6.00 per hour. This is 50 cents less per hour than workers at Wendy's. If you work for 8 hours, how much will you earn at Wendy's?" Scores could range from 0–12.

Procedure

Study 1 was completed online via Qualtrics. After consent was obtained, participants completed a series of demographic questions, they then completed the CRT and constraint task, and finally the numeracy test. Prior to the CRT and constraint task, participants were provided with general instructions. Further instructions were presented on subsequent pages as appropriate, for example when the matrix pattern appeared, participants were instructed: "Remember this grid pattern, you will be asked to recall it later." The same materials were presented to every participant. However, the order in which the materials were presented varied between the no-load and load conditions. All participants were first presented with the information component of the question. The information component of the questions was presented first, before the matrix pattern, to minimise the effect of the load on the comprehension process (Van Lier et al., 2013). For example, the information component of the bat and ball problem stated: "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball." All participants were then presented with the matrix pattern to memorise. From this point, the participants in the no-load condition proceeded directly to the 'recall' page where they clicked the squares corresponding to the preceding matrix. No-load participants were then presented with the full CRT item to solve. In contrast, after memorising the matrix pattern, participants in the load condition were then asked to solve the CRT item (while remembering the matrix pattern), and—finally—to recall the matrix pattern (see Table 1). Those in the load condition were therefore required to solve the problems under cognitive constraint, whereas those in the no-load condition were able to solve the CRT item with no simultaneous task or constraint.

Results

Preliminary analysis

Numeracy was analysed to check the categorisation of mathematical experience by course and occupation. High experience participants scored higher ($M = 10.27$, $SD = 1.53$) than intermediate experience participants ($M = 9.70$, $SD = 1.79$) who, in turn, scored higher than low experience participants ($M = 9.35$, $SD = 1.94$). These differences were not significant, $F(2, 61) = 1.24$, $p = .295$, $\eta_p^2 = .04$, likely due to ceiling effects. However, the positive trend between experience categories and numeracy supports the grouping of participants by course and occupation as a measure of mathematical experience. Additionally, that all participants performed well on the numeracy test suggests that differences in CRT performance were not due to a general lack of mathematical ability by any group.

Performance on the matrix memory constraint was analysed to check for systematic differences in task preference between the experience groups; that is, whether participants prioritised the matrix task over the CRT task, or vice versa. The effect of experience and constraint condition on matrix performance was assessed using a two-way ANOVA. Participants in the no-load condition ($M = 92.05$, $SD = 14.39$) outperformed those in the load condition ($M = 80.52$, $SD = 11.67$) on the matrix memory task, $F(1, 59)=16.16$, $p<.001$, $\eta_p^2=.22$. The main effect of experience on matrix performance was not significant, $F(2, 59)=.74$, $p=.480$, $\eta_p^2=.03$. There was a significant interaction between experience and condition on matrix performance: $F(2, 59)=3.29$, $p=.044$, $\eta_p^2=.10$. This indicated that the effect of constraint condition on matrix performance was different at different levels of experience.

An interaction contrast test was run comparing matrix performance for those in the low and high experience groups combined, compared to those in the intermediate experience group. This contrast revealed that the effect of load condition on matrix performance was greater for intermediate participants than low and high experience participants, $F(1, 59)=6.52$, $p=.013$, $\eta_p^2=.10$. This indicated that intermediate participants may have prioritised the CRT over the matrix task to a greater extent than did low and high experience participants. Conversely, low and high experience participants may have prioritised the matrix task over the CRT more than intermediate experience participants. This pattern of results suggested a potential for different task preference between the low and high experience groups and the intermediate experience group in line with the primary hypotheses. Therefore, we included matrix performance as a potential covariate in the main analyses.

Main analysis

To test our three hypotheses, we used a 2 (constraint) x 3 (experience) between-subjects ANOVA, with pairwise comparisons to follow up main-effects, and interaction contrasts to follow up interaction effects. CRT scores were significantly higher for participants with greater experience, averaged across load condition, $F(2, 59)=31.02$, $p<.001$, $\eta_p^2=.51$. High experience participants ($M = 2.60$, $SD=.63$) outperformed intermediate experience participants ($M = 1.54$, $SD = 1.06$), $F(1, 64)=14.76$, $p<.001$, $\eta_p^2=.19$, who outperformed low experience participants ($M=.65$, $SD=.69$), $F(1, 64)=14.04$, $p<.001$, $\eta_p^2=.19$. The main effect of load was not significant when averaged across experience, $F(1, 59)=2.04$, $p=.158$, $\eta_p^2=.03$. However, there was a significant interaction between experience and constraint condition on CRT performance, $F(2, 59)=6.66$, $p=.002$, $\eta_p^2=.18$.

An interaction contrast test was run to compare the difference in CRT scores between the load and no-load constraint conditions for the low and high experience groups combined, compared to the intermediate experience group. Results revealed that the effect of load constraint was greater for intermediates than for the low and high experience groups,

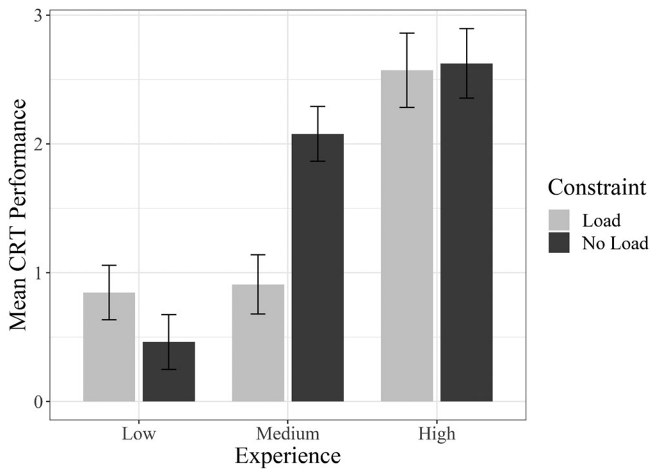


Figure 2. CRT performance by level of experience and constraint. Error bars reflect +/- 1 SE.

$F(1, 59)=11.18$, $p=.001$, $\eta_p^2=.16$. These results are presented in Figure 2. Due to differences between groups on matrix performance (reported above) we also ran this analysis with matrix performance included as a potential covariate, however, the pattern of results was unchanged[2]. Therefore, we have reported the results excluding matrix performance.

Discussion

Study 1 examined three hypotheses. In support of our first hypothesis, CRT performance was higher for those with greater mathematical experience. This finding is in line with previous studies that have shown positive associations between CRT performance and mathematical factors such as numeracy (Sinayev & Peters, 2015; Welsh et al., 2013) and SAT scores (Frederick, 2005; Obrecht et al., 2009; Thompson et al., 2013). Our second hypothesis was not supported, CRT performance was not affected by constraint when averaged across experience. This is similar to Bago and De Neys' (2019) finding that participants were able to complete the bat-and-ball problem under cognitive constraint but contrasts with Johnson et al. (2014, 2016) who observed a detrimental effect of load on CRT performance. However, as the results pertaining to our third hypothesis indicate, the effect of constraint should be considered in relation to experience.

In support of our third hypothesis, load constraint and experience interacted to affect CRT performance. The performance of intermediate participants was affected by the constraint to a greater extent than the

[2]These results are available from the authors on request.

performance of low or high experience participants. This indicates that intermediate participants were able to reach the correct solutions on the CRT when sufficient working memory resources were available; they were able to reach the correct solutions via Type 2 thinking. The high experience group, conversely, did not show lowered performance under constraint; they were able to reach the correct solutions via Type 1 thinking.

The nature of reasoning for the low experience group is less clear. The constraint manipulation did not affect the performance of low experience participants which may indicate that they had engaged Type 1 thinking. However, it is also possible that the group was employing Type 2 thinking but that the deliberation was ineffective in producing the correct solution possibly due to a lack of mindware (Stanovich, 2018) or, for example, less analytic dispositions (Pennycook et al., 2015). Although the low performance in both load and no load conditions was expected and in line with Stanovich (2018) and De Neys' (2012, 2014) models, future studies will be needed to disentangle whether those with low experience were engaged in Type 1 processing or if they were engaging in ineffective Type 2 processing. One way to clarify which Type of reasoning was employed by participants exhibiting poor performance is to look at response times in conjunction with accuracy. Slow responses and low performance would be indicative of ineffective Type 2 processing, whereas, fast responses and low performance would be indicative of Type 1 processing (see Stanovich, 2018; Pennycook et al., 2015). While there are remaining questions about the nature of the thinking for low experience participants, the pattern of results for the intermediate and high experience groups provides reasonable evidence to suggest that these groups had employed Type 2 and Type 1 processing, respectively.

The use of groups differing in real-world mathematical experience increased the study's ecological validity but also introduced limitations like confounds and a limited sample size. Between-group differences in demographic factors like gender and age are likely to occur when employing between-subject experience manipulations. In Study 1, gender was prevented from becoming a potential confound by including an all-male sample, however, systematic differences in age were observed. Among adults, aging is related to declines in executive functioning which could increase the detrimental impact of cognitive constraints (Buckner, 2004; Park et al., 1996). In Study 1, the high experience participants were impacted by the load to a lesser extent than the younger intermediate group. Therefore, the age-related differences are unlikely to account for the current findings. The inclusion of an expert, high experience population, lead to a small sample size which could limit the generalisability of the findings. The use of preexisting indicators of experience is likely to bring demographic disparities mirroring those in society and restrict sample size. While it is important to examine the role of real-world differences in experience in reasoning, these issues highlight the need for a within-subject experience manipulation.

In addition to demographic factors, the experience groups may have differed in their prior exposure to the CRT or psychological factors. Bialek and Pennycook (2018) found that exposure to the CRT did not impact performance, and Janssen et al. (2020) found that even with accuracy feedback (correct/incorrect) there was no significant improvement in performance. This said, we cannot be sure if these factors influenced the Study 1 findings. One way to reduce the likelihood of exposure confounds is to couple a within-subject experience paradigm with random allocation to constraint conditions. Additionally, experience groups may have differed on factors, such as thinking dispositions or WMC, that have demonstrated associations with CRT performance (e.g., Toplak et al., 2011). WMC is a particularly important factor for consideration given the current use of a cognitive load manipulation. It could be that high experience participants had larger WMCs than intermediate and low experience participants. Consequently, the high experience participants may have been able to engage in Type 2 processing even with the additional cognitive load. That is, the load manipulation—while successfully used in previous studies—may not have been large enough to “knock out” the WMC of those in the high experience group. These issues are directly addressed in Study 2, which included a within-subject manipulation of experience with an additional, harder constraint and a measure of WMC.

Study 1 is the first experiment to combine experience and cognitive constraint manipulations in a study of CRT performance. The findings showed reasonable support for the dual process assertion that Type 2 processes can become Type 1 processes with increases in experience. It also supports the logical intuition model’s assertion that logico-mathematical principles can be enacted via Type 1 processes (De Neys, 2012). In line with the interpretations of previous cognitive constraint studies and dual process theory (e.g., Bago & De Neys, 2019; Stanovich, 2018), the results from Study 1 suggest tentatively that low experience participants were using Type 1 processes, and more convincingly that intermediate and high experience participants were using Type 2 and Type 1 processing, respectively. These findings are largely consistent with the hypothesised relationship between experience and thinking Type.

Study 2

As in Study 1, Study 2 aimed to examine the relationship between experience and thinking Type. Study 2 also aimed to address the limitations in

Study 1 and develop a more sustainable method for the inclusion of experience in future studies. This was achieved with a within-subject training manipulation of mathematical experience, the inclusion of a more taxing cognitive constraint, the consideration of individual differences and the examination of response times. This approach yielded a 3 (constraint) x 3 (test point) x WMC (continuous predictor) mixed design. As in Study 1, we manipulated cognitive constraint with matrix tasks, in this case with three levels: low, medium, and high. We examined CRT performance and response times at three test points: before training (T1), halfway through training (T2), and after training (T3).

We predicted that WMC would influence performance due to the relationships between WMC and cognitive load, and between WMC and automation (via experience). Holding the point of automation constant, the greater an individual's WMC, the less impact a cognitive load was expected to have on performance (e.g., Baddeley, 1986; Engle et al., 1999; Hambrick & Engle, 2002). WMC also affects speed of learning and subsequently, the level of experience at which an individual will reach automation (e.g., Baddeley, 1986; Kyllonen, 1996; Kyllonen & Stephens, 1990). Individuals with greater WMC were expected to automate processes more quickly and "free up" working memory resources such that the reasoner would have a greater capacity to withstand cognitive constraints as well as a greater capacity for further learning and automation. These relationships were expected to form a dynamic and cumulative effect of WMC on performance across training. As such we expected a three-way interaction between WMC, test point and constraint on CRT performance.

In Study 2 we also included an examination of response times. In line with the mindware continuum, we expected that response times would be fast at T1 (prior to training) due to the reasoner not having the adequate mindware to detect conflict and engage Type 2 processes. In line with the mindware continuum and the results in Study 1, we expected that response times would increase at T2, reflecting Type 2 processing and the "Zone of Conflict", and decrease at T3, reflecting a shift back to Type 1 processing. At this point, we should note that the extent of training needed to automate the processes required for completing the CRT was unknown, hence, we put forward these hypotheses tentatively regarding specific test points, but strongly regarding the expected pattern.

In sum, we expected that:

1. Performance would increase with increases in experience (test point);
2. Test point, constraint, and WMC would interact to affect performance such that at T2, participants with lower WMC would be negatively affected by the load manipulation to a greater extent than those with

higher WMC; this difference was expected to be smaller at T1 and T3 than at T2.

3. Response times were expected to increase from T1 to T2 and decrease from T2 to T3.

Method

Participants and design

A total of 85 participants were recruited[3]. Participants were undergraduate psychology students at Macquarie University, Sydney (22 males, 61 females, 2 unspecified). Ages ranged from 17 to 43 ($M = 20.28$, $SD = 4.00$). Participants were awarded course credit for participation. Study 2 used a 3 (experience: within-subject) x 3 (constraint: between-subject) x WMC (continuous predictor) mixed design. Experience was included as a within-subject factor with three levels: T1, T2, and T3. Constraint was included as a between-subjects factor; participants were randomly allocated to one of three conditions: low load (N = 27), medium load (N = 30), or high load (N = 28).

Materials

CRT. The CRT was presented to the participants at each test point (T1, T2, and T3; Frederick, 2005). The CRT questions were not included in the training sections and the participants were never exposed to the solutions. Although the participants would see the CRT questions at three points in the study, we note that this repetition, alone, was not expected to elicit increased performance. A recent study by Raoelison and De Neys (2019) presented participants with the bat and ball problem 50 times and found the majority of participants continued to demonstrate the same pattern of responding from start to finish. Therefore, rather than risk inadvertently changing the difficulty or nature of the scale between test points, the participants received the same items. As in Study 1, a free-response format was used. Scores could range from 0–3 at each test point.

Cognitive constraint. As in Study 1, matrix memory tasks were included as cognitive constraints. Study 2 employed low, medium, and high constraint conditions. Low constraint matrices had three coloured squares in a 3×3 grid that formed horizontal, vertical or diagonal lines. The medium constraint matrices had four coloured squares in 3×3 grids that formed three-

[3]The sample size was based on detecting similar effects sizes as those observed in Study 1. Study 2 was expected to yield greater power because of the within-subject manipulation of experience; however, it also included an additional variable, WMC. Taking these factors into account, a larger number of participants were recruited. A power analysis conducted in R with 1000 simulations revealed that this design would detect moderate to large effects with N = 84, 86.50% (CI: 84.22, 88.56).

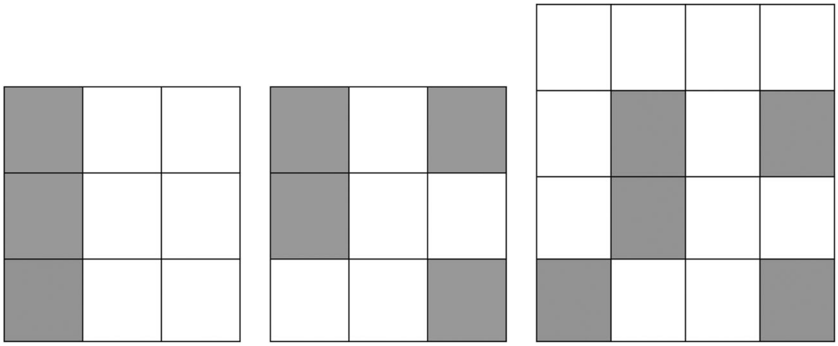


Figure 3. Examples of matrices used for (A) low, (B) medium, and (C) high constraint conditions.

piece patterns. The high constraint matrices had five coloured squares in 4×4 grids (e.g., Johnson et al., 2016; Trémolière et al., 2012). See Figure 3.

Training materials. Eighteen training items were developed to reflect the structure of the original CRT items. To avoid rote-learning, training items included different numbers and content to the original CRT items. Six training items were created per CRT item. An example of an original CRT item is: "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?" An example of a corresponding training item is: "A pen and a notebook cost $25 in total. The notebook costs $5 more than the pen. How much does the pen cost?" Unlike the test items, training items included feedback (correct/incorrect) and guidance to help participants reach the correct solution.

The feedback was tailored to the solution processes that each problem-structure required. The bat and ball item, for example, can be solved using algebra and substitution. Hence, for the six training items pertaining to the bat and ball problem, participants who gave the incorrect response were guided through a process of breaking the problem down into algebra and solving it via substitution. The problem-structure of the CRT items 2 and 3 do not have simple algebraic solutions (for example, item 3 would require an understanding of exponential growth) but they do lend themselves to worded explanations. Therefore, feedback for items reflecting problems 2 and 3 of the CRT was provided as a written explanation. See Supplementary Material, Table 1, for examples of feedback for each of the three problem-structures.

Numeracy. Numeracy was assessed using the Berlin Numeracy Scale (Cokely et al., 2012). The scale employed accuracy-based conditional branching and, therefore, contained included two to four items. For example, if a participant gave an incorrect response, they were then presented with an easier question. The scale was developed for highly educated samples such as college students and has been used in conjunction with the CRT in several recent studies (Primi et al., 2018). An example is "Out of 1,000 people in a small town, 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir?" Scores had a possible range of 1 to 4.

WMC. WMC was assessed using the short Operation Span Task (Foster et al., 2015; Unsworth et al., 2005). The task required participants to remember and then recall a list of letters while intermittently assessing the validity of several short mathematical problems. A two-letter task, for example, would start with an equation (e.g., "(9/3) − 2 = 2?") and the respondent must select True or False. Following this, the respondent is presented with a letter (e.g., "D".) Next, they are presented with another equation to assess (e.g., "(8/4) − 1 = 1?"), followed by another letter (e.g., "E".) A recall sheet is then provided so that the respondent can serially recall the letters they have seen. For this example, the respondent would list D, and then E. The span task in Study 2 included 3 practice trials with letter spans of 2, and then six test trials with letter-spans of 2 to 7. Accurate serial recall was summed; scores had a possible range of 0 to 27.

Procedure

Testing was completed in person, in groups of five to eight. Participants completed the study on computers in partitioned booths and were provided with a notepad and pen. Once consent was obtained, participants completed a series of demographic questions. They were then given instructions for the general procedure, including instructions that the notepad could be used to assist with mathematical working out but not for the memory task components: matrix patterns or letter sequences. Notepads were checked to ensure participants did not use them for the memory tasks. Instructions were also presented with each question as appropriate, for example: "Submit your final answer only. Use numbers only (up to 2 decimal places). Exclude symbols or words e.g. $, cents, km." The numeracy test was completed first, then, in random order, the WMC test and the CRT training task were completed.

The training task included five blocks: test 1 (T1), training 1, test 2 (T2), training 2, and test 3 (T3). Each test block included the three original CRT items, no solutions were provided. The two training blocks included nine items each, three per problem structure (see Materials). Test and training items were presented in the same order for all participants. All participants completed the test problems whilst remembering a matrix pattern. However, the complexity of the matrices was dependent on the participant's constraint condition: low, medium, or high load. At T1, as in Study 1, the information component was presented, then the matrix pattern to memorise (800 ms), the full question, and then a blank grid for recalling the matrix pattern (see Table 1). At T2 and T3 the information component was not presented prior to the matrix pattern. This was to prevent the participants recognising the problem, predicting the question, and working on it before the load was enforced. At T2 and T3, a matrix pattern was presented (800 ms), then the full question until a response was made, and finally a blank grid for the participant to recall the matrix pattern until a response was made. Different matrix patterns were used for each item and each test point. To ensure that any training effects would not be confounded by the order of matrix patterns, the order of matrix patterns was counterbalanced such that half of the participants received matrix patterns 1 to 9 and the other half received the same matrix patterns but in reverse order 9 to 1.

Results

Preliminary analysis

To ensure that only the participants who put genuine effort into the constraint manipulation were included, those with overall matrix scores more than two standard deviations below the mean for their condition were excluded. Subsequently, five participants were excluded: Two from the medium constraint condition and three from the high constraint condition. Eighty participants remained in the final analyses (N_{low}=27, N_{medium}=28, N_{high}=25). The training paradigm was highly effective. Participants' CRT performance increased from T1 ($M = 1.10$, $SD = 1.06$) to T2 ($M = 2.43$, SD=.91), and from T2 to T3 ($M = 2.57$, SD=.79). Participants demonstrated reasonable numeracy ($M = 2.53$, $SD = 1.18$) and WMC ($M = 20.46$, $SD = 5.11$). The correlation between numeracy and WMC was positive and marginally significant $r(33)$=.319, p=.051[4]. Before assessing our hypotheses, we examined whether numeracy, WMC, and CRT performance at T1 differed as a function of constraint condition. One-way between-subjects ANOVAs did not reveal any significant differences between constraint conditions on numeracy, $F(2, 38)=1.91$, p=.164, η_p^2=.098, WMC $F(2, 77)$=.824, p=.442, η_p^2=.021, or CRT performance at T1, $F(2, 77)$=.405, p=.668, η_p^2=.010.

[4]Some data was missing for this analysis due to a technical issue with the numeracy scale.

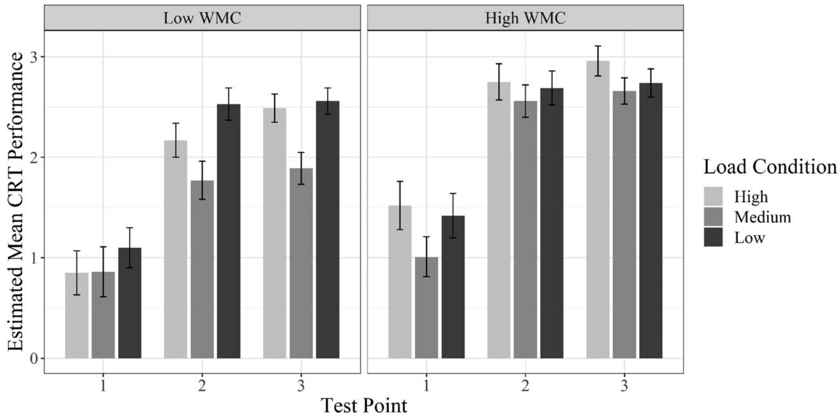


Figure 4. Estimated means for CRT performance by test point and constraint and with the model adjusted to low WMC (left panel) and high WMC (right panel). Error bars +/- 1 SE.

Main analysis

To test hypotheses (1) and (2), a general linear model was employed with three predictors: experience (3 levels; within-subjects factor), constraint (3 levels; between-subjects factor), and WMC (continuous predictor). The model revealed significant main effects of experience, $F(1.32, 97.89)=124.09$, $p<.001$, $\eta_p^2=.626$, and WMC, $F(1, 74)=35.73$, $p<.001$, $\eta_p^2=.326$, on CRT performance. Additionally, a significant three-way interaction was observed between experience, constraint and WMC on CRT performance, $F(2.65, 97.89)=3.81$, $p=.016$, $\eta_p^2=.093$. No other effects reached statistical significance (see Supplementary Material, Table 2). To examine the three-way interaction, we analysed the two-way interactions between experience and constraint on CRT performance, with the model adjusted to low WMC (-.5 SD) and then with the model adjusted to high WMC (+.5 SD). These results are presented in Figure 4.

Low WMC. When the model was evaluated at low WMC, the effect of load was not significant at T1, $F(2,74)=.44$, $p=.645$, $\eta_p^2=.012$. However, it was significant at T2, $F(2,74)=4.83$, $p=.004$, $\eta_p^2=.115$, and T3, $F(2,74)=5.81$, $p=.005$, $\eta_p^2=.136$. Tests of simple effects revealed that at T2 the mean performance for participants in the low constraint condition was greater than that of those in the medium constraint condition, $F(1, 74)=9.57$, $p=.003$, $\eta_p^2=0.115$, but no different from that for the high constraint condition, $F(1,74)=2.54$, $p=.115$, $\eta_p^2=.033$. The mean performance for participants in the medium constraint condition was no different to that of those in the high constraint condition, $F(1,74)=2.46$, $p=.121$, $\eta_p^2=.032$. At T3 the mean performance for participants in the medium constraint condition was lower than both the low constraint condition, $F(1, 74)=10.25$, $p=.002$, $\eta_p^2=.122$, and the high constraint condition, $F(1, 74)=7.91$, $p=.006$, $\eta_p^2=.097$. As for T2, there was no significant difference between the low constraint condition and the high constraint condition, $F(1,74)=.11$, $p=.739$, $\eta_p^2=.001$.

Although there was, as expected, a three-way interaction between experience, constraint and WMC, the pattern of the effect of constraint only partly followed the expected pattern. In contrast to our hypotheses, the high constraint condition yielded higher performance than the medium constraint condition. One explanation for this, is that the constraint may have impacted response times rather than accuracy[5]. To examine this possibility, we tested the effect of the three-way interaction on response times. The model revealed a main effect of test point, $F(1.28, 79.16)=38.07$, $p<.001$, $\eta_p^2=.380$. Response time (sec) significantly decreased from T1 ($M=85.82$, $SE=7.04$) to T2 ($M=56.89$, $SE=3.90$), $F(1,61)=$, $p<.001$, $\eta_p^2=.206$, and from T2 to T3 ($M=32.77$, $SE=2.13$), $F(1,61)=63.37$, $p<.001$, $\eta_p^2=.509$. However, no other effects reached statistical significance suggesting that the accuracy of high constraint participants was not offset by longer response times (see Supplementary Material, Table 3).

High WMC. In contrast to low WMC, when the model was evaluated at high WMC, the effect of constraint was not significant at T1, $F(2,74)=1.594$, $p=.210$, $\eta_p^2=.041$, T2, $F(2,74)=.324$, $p=.724$, $\eta_p^2=.009$, or T3, $F(2,74)=1.098$, $p=.339$, $\eta_p^2=.029$.

Hypothesis 3 predicted that response times would increase from T1 to T2 and decrease from T2 to T3. To test this hypothesis, we ran a one-way repeated ANOVA with response time as the dependent variable and experience (3 levels: T1, T2, T3) as the independent variable. Four participants were removed from this analysis due to having response times greater than 3 SD from the mean at any one test point. For the remaining participants ($N=76$), there was a significant effect of experience on response time, $F(1.25, 93.86)=26.75$, $p<.001$, $\eta_p^2=.263$. Counter to our predictions, contrast tests revealed that response time (sec) decreased from T1 ($M=90.49$, $SE=7.74$) to T2 ($M=64.96$, $SE=4.56$), $F(1,75)=8.39$, $p=.005$, $\eta_p^2=.101$, and from T2 to T3 ($M=38.02$, $SE=2.89$), $F(1,75)=58.37$, $p<.001$, $\eta_p^2=.438$. We observed, however, that the variance in response times was greater at T1 than T2 and T3. This large variability in response times at T1 suggested that some of the participants may have been employing faster Type 1 responses while others were employing slower Type 2 responses, resulting in a large spread of response times.

[5]We would like to thank our reviewer, Valerie Thompson, for this suggestion.

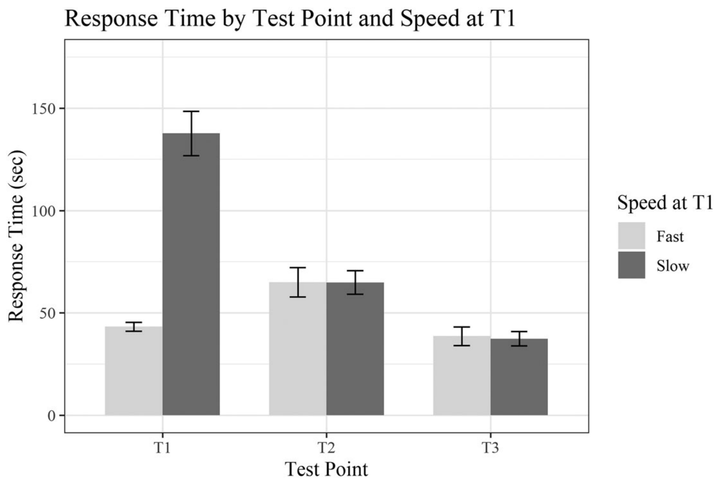


Figure 5. Mean response times for each participant by test point and speed at T1. Colours reflect the median split of participants by response times at T1. Error bars +/-1 SE.

To explore this possibility, we divided the participants using a median split based on response times at T1 (72.36 sec) and examined the effect of experience separately for each group. For those with response times lower than 72.36 sec at T1, experience had an overall effect on response time, $F(1,75)=58.37$, $p<.001$, $\eta_p^2=.438$. In line with our initial hypotheses, their response times significantly increased from T1 ($M=43.26$, $SE=2.21$) to T2 ($M=65.05$, $SE=7.16$), $F(1,36)=8.76$, $p=.005$, $\eta_p^2=.196$, and decreased from T2 to T3 ($M=38.62$, $SE=4.59$), $F(1,36)=30.41$, $p<.001$, $\eta_p^2=.458$. For those with response times higher than 72.36 sec at T1, experience also had an overall effect, $F(1.31,48.41)=58.04$, $p<.001$, $\eta_p^2=.611$. Their response times significantly decreased from T1 ($M=137.73$, $SE=10.82$) to T2($M=64.87$, $SE=5.76$, $F(1,37)=37.95$, $p<.001$, $\eta_p^2=.506$, and from T2 to T3 ($M=37.42$, $SE=3.58$), $F(1,37)=27.48$, $p<.001$, $\eta_p^2=.426$. These findings are presented in Figure 5.

Discussion

Study 2 examined three hypotheses. The first hypothesis, that CRT performance would increase with training was supported. The second hypothesis, that a three-way interaction between test point, WMC and constraint would impact performance was partially supported. The third hypothesis, that response time would increase from T1 to T2 and decrease from T2 to T3 was also partially supported. In line with the second hypothesis, test point, WMC, and constraint interacted to affect performance. At T1, the high WMC participants showed slightly higher performance than the low WMC participants but did not show an effect of constraint. This suggests that some participants with high WMC may have been resilient to the effect of load, even prior to training. In addition to having the cognitive hardware to sustain greater cognitive loads, these participants may have had higher numeracy (i.e. more advanced mindware) than their low WMC counterparts. This is supported by the positive trend between WMC and numeracy in the current study and by previous studies that have established an association between these factors (e.g, Cokely et al., 2012).

At T2 and T3, the high WMC participants were able to successfully complete the problems, even under constraint. This could be due to the relationship between WMC and cognitive load. That is, a capacity effect may have occurred such that greater WMC reduced the impediment caused by the cognitive load because the participant had sufficient WMC to perform both tasks. Alternatively, it could be because of the relationship between WMC and automation. That is, a learning effect may have occurred such that higher WMC led to faster transitions from Type 2 to Type 1. Alternatively, a combination of capacity and learning effects may have led to the high WMC participants' strong performance under constraint at T2 and T3. Overall, the high WMC participants appear to have learned how to solve the problems, as indicated by their improved performance but, additionally, exhibited a transition to Type 1, as indicated by their resilience under constraint.

In contrast, the low WMC participants showed improved performance but less resilience to the constraints. For participants with low WMC, load had a detrimental effect on performance at T2 and T3 but not at T1. Broadly, this suggests that the training was effective in teaching the low WMC participants how to solve the problems, but was not sufficient for automating the solution process. A closer examination of the effect of constraint at low WMC revealed that, as expected, the performance of participants in the low constraint condition was higher than that of those in the medium constraint condition but, in contrast to our expectations, equal to that of the high constraint condition. This is somewhat consistent with Study 1 in that the four-piece matrices were effective in reducing performance in both Study 1 and Study 2. The ineffectiveness of the five-piece matrices employed in Study 2 may reflect a methodological issue in that the higher constraint matrices may not have been as taxing as the medium constraints. However, this seems unlikely given the success of this constraint in previous studies (e.g., Trémolière et al., 2012). Another possibility is that the high constraint group included people who, despite lower WMC, were still able to circumvent the constraint effects. This introduces the possibility that confounding factors may need to be considered.

In the current study, numeracy and initial success on the CRT (performance at T1) did not differ significantly between the constraint groups and are therefore unlikely to account for the difference in performance of the high and medium constraint groups. However, and despite random allocation of participants to conditions, it is possible that the groups could have differed on other factors that could influence automation such as IQ, thinking dispositions, self-efficacy or motivation. This pattern of results highlights a particularly pressing question for future research: Which individual differences affect the automation continuum, and how? Although the performance of the high constraint group deviated from the expected pattern of performance, the remaining results from Study 2 are generally consistent with the proposed relationship between experience and reasoning Type.

The third hypothesis was related to proposed changes in response times across time points. Overall, response times reduced significantly from T1 to T2 and T2 to T3. However, a closer examination revealed that some participants followed the expected pattern of results; response times increased from T1 to T2 and decreased from T2 to T3. In line with our hypothesis, this indicates a transition from Type 1, to Type 2, and back to Type 1 processing. In contrast, others displayed a pattern of consistently decreasing response times which indicates that they may have employed ineffective (i.e. low accuracy) Type 2 processing at T1, effective Type 2 processing at T2 and effective Type 1 processing at T3. Therefore, although most participants exhibited poor performance at T1 it seems that at least some were engaged in slower Type 2 processes. Interestingly, despite the large variance observed for response times at T1, the participants' response times converged at T2 and T3. This is a preliminary indication that, prior to training, some people are less likely to engage in deliberative thinking than others. After just a small amount of practice however, the difference between these groups appears negligible.

Response times indicated that some participants followed the expected pattern of changes in thinking Type across training, while others did not. An explanation for these different trajectories is offered by dual process models that incorporate the possibility of ineffective Type 2 processing. For example, Pennycook et al. (2015) three-stage dual process model suggests that conflict detection leads to Type 2 processing but that conflict detection, alone, is not sufficient for successfully overriding incorrect Type 1 responses. His idea is extended by Stanovich (2018) who postulated a need for a 2×2 classification system using response times and accuracy. For those giving incorrect responses, as in the current Study at T1, fast responding indicates inaccuracy due to a lack of mindware or failure to detect conflict, whereas, slow responding indicated inaccuracy due to a failure to override the incorrect default response. It may be that the two groups (fast and slow responders) in the current study fall into these categories. However, the current findings must be replicated and extended to gain a deeper understanding of the individual and cognitive characteristics that distinguish these groups and their unique trajectories of automation.

General discussion

In two studies, we used manipulations of experience and cognitive constraints to examine the relationship between experience and thinking Type. Generally, the results supported the claim that Type 2 processes can transition into Type 1 processes with practice. In Study 1, we found evidence that participants with different levels of real-world mathematical experience, as operationalised via university degree and occupation, showed differences in performance and thinking Type when completing the CRT. In Study 2, we found evidence that as participants' mathematical training increased, they exhibited increased performance, changes in response times and shifts in thinking Type. Together, these studies suggest that real-world changes in experience are associated with differences in thinking Types and that these differences can be generated in experimental settings. Overall, the findings support the dual process' automation hypothesis—that Type 2 processes can become Type 1 processes over time (De Neys & Pennycook, 2019; Stanovich, 2018). Additionally, the findings support the suggestion that previously varied findings may be accounted for by individual differences in experience (Bago & De Neys, 2019) and that the mechanisms underlying correct responding on the CRT may be mediated by experience.

Previous studies have employed cognitive constraints to test which Type of thinking is engaged when people are solving the CRT. Some have found support for the assertion that correct solutions to the CRT require Type 2 processing (Johnson et al., 2016) while others have found evidence that correct solutions can also be reached via Type 1 processing (Bago & De Neys, 2019). In line with previous commentary (Bago & De Neys, 2019), the present findings suggest that mathematical experience may, at least in part, account for this variability. The studies that have shown a detrimental effect of cognitive constraint on CRT performance may have included participants at an intermediate phase of experience. That is, participants who were able to reach the solution via Type 2 processes. In contrast, studies in which constraints had a smaller effect on performance may have included participants with greater mathematical experience. That is, participants who could reach the correct solution via Type 1 processes. In addition to accounting for this variability in previous findings, the role of experience may also facilitate the integration of competing theories about the mechanisms underlying Type 1 and Type 2 processing.

Traditional explanations for performance on bias tasks suggested that correct responding was the result of Type 2 dependent, default-intervention processes (e.g., for the CRT: Kahneman, 2011; Kahneman & Frederick, 2005). The default-intervention position asserts that incorrect default responses generated via Type 1 processes can be overridden and corrected by Type 2 processes (Evans, 2006; Evans & Stanovich, 2013). However, recent instances of correct responding by participants under cognitive constraint has been presented as a challenge to the default-intervention explanation for correct responding on bias tasks (e.g., Bago & De Neys, 2017, 2019). For example, studies using two-response paradigms have found that some participants who gave correct responses under Type 2 conditions were also able to give correct responses under Type 1 conditions (e.g., Pennycook & Thompson, 2012; Thompson & Johnson, 2014; Thompson et al., 2011). Considered alongside the current findings it seems likely that Type 2 dependent, default-intervention processes do not underlie correct responding by people with high levels of relevant experience. However, default-intervention processes may underlie correct responding by people with intermediate levels of experience. Future studies could investigate whether the processes underlying correct responses are mediated by experience, for example, by combining experience manipulations with a two-response paradigm.

The present findings can be accounted for by Stanovich's (2018) mindware continuum. The findings from Study one reflect a parabolic relationship between experiment and thinking Type—from Type 1 to Type 2 and back to Type 1—however, a potential floor effect left some uncertainty regarding the nature of reasoning by low experience participants. This was addressed in Study 2 by incorporating an analysis of response times. Study 2 demonstrated that for some people there was, as hypothesised, a parabolic learning trajectory reflecting the full mindware continuum. In these cases, low experience was associated with fast but incorrect responses (i.e. Type 1), intermediate was associated with slower and cognitively demanding but correct responses (i.e. Type 2), and high experience was associated with fast but correct responses (i.e. Type 1).

However, for the remaining participants, an unexpected trajectory was observed. Low experience was associated with slower and incorrect responding (i.e. ineffective Type 2), intermediate was associated with effective Type 2 processing, and high experience with Type 1. This unexpected pattern can be accounted for within the mindware continuum provided these participants were at an intermediate stage of experience prior to training. However, that is not necessarily the case, for example, it may be that regardless of experience, some individuals were more likely to engage in deliberative thinking than others. These findings highlight two important queries for future research: First, who is more likely to engage in deliberative thinking at novice stages of experience and what are the underlying processes behind that engagement? Second, do differences in deliberation influence an individual's subsequent learning trajectory and, if so, how?

The transition of reasoning processes from effortful to easy is a commonly assumed but rarely investigated concept. This article presents a unique combination of experience manipulations and cognitive constraints to explore the relationship between experience and reasoning. It provides the first explicit empirical support for the dual process assumption that Type 2 processes can become Type 1 processes with practice and, in doing so, brings to light new and important queries for the continued examination of the relationship between experience and reasoning.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Baddeley, A. (1986). *Working memory*. Clarendon Press.

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90–109. https://doi.org/10.1016/j.cognition.2016.10.014

Bago, B., & De Neys, W. (2019). The Smart System 1: evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *25*(3), 257–299. https://doi.org/10.1080/13546783.2018.1507949

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 12–23. https://doi.org/10.1037/0096-1523.14.1.12

Bialek, M., & Pennycook, G. (2018). The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods*, *50*(5), 1953–1959. https://doi.org/10.3758/s13428-017-0963-x

Buckner, R. L. (2004). Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron*, *44*(1), 195–208. https://doi.org/10.1016/j.neuron.2004.09.006

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The berlin numeracy test. *Judgment and Decision Making*, *7*(1), 25–47. http://simsrad.net.ocs.mq.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=71325917&site=ehost-live

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*(1), 20–33.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science : a Journal of the Association for Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*(2), 169–187. https://doi.org/10.1080/13546783.2013.854725

De Neys, W. (2017). *Dual Process Theory 2.0*. Routledge. https://doi.org/10.4324/9781315204550

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177/0963721419855658

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331. https://doi.org/10.1037//0096-3445.128.3.309

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *The American Psychologist*, *49*(8), 709–724. https://doi.org/10.1037/0003-066X.49.8.709

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395. https://doi.org/10.3758/bf03193858

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. https://doi.org/10.3758/s13421-014-0461-7

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, *44*(4), 339–387. https://doi.org/10.1006/cogp.2001.0769

Handley, S. J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new parallel processing model. *The Psychology of Learning and Motivation*. *62*, 33–58. https://doi.org/10.1016/bs.plm.2014.09.002

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, *87*(1), 18–32. https://doi.org/10.1037/0022-0663.87.1.18

Janssen, E. M., Raoelison, M., & De Neys, W. (2020). "You're wrong!": The impact of accuracy feedback on the bat-and-ball problem!. *Acta Psychologica*, *206*, 103042. https://doi.org/10.1016/j.actpsy.2020.103042

Johnson, E. D., Tubau, E., De Neys, W. (2014). The unbearable burden of executive load on cognitive reflection: A validation of dual process theory. *Proceedings of the Annual Meeting of the Cognitive Science Society, 36*, 2441–2446.

Johnson, E. D., Tubau, E., & De Neys, W. (2016). The doubting system 1: Evidence for automatic substitution sensitivity. *Acta Psychologica*, *164*, 56–64. https://doi.org/10.1016/j.actpsy.2015.12.008

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge University Press.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49–75). Lawrence Erlbaum Associates, Inc.

Kyllonen, P. C., & Stephens, D. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences*, *2*(2), 129–160. https://doi.org/10.1016/1041-6080(90)90020-H

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, *25*(4), 361–381. https://doi.org/10.1002/bdm.752

Morsanyi, K., Prado, J., & Richland, L. E. (2018). Editorial: The role of reasoning in mathematical thinking. *Thinking & Reasoning*, *24*(2), 129–137. https://doi.org/10.1080/13546783.2018.1435425

Obrecht, N. A., Chapman, G. B., & Gelman, R. (2009). An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition*, *37*(5), 632–643. https://doi.org/10.3758/MC.37.5.632

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging*, *11*(4), 621–637. https://doi.org/10.1037/0882-7974.11.4.621

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., & Thompson, V. A. (2012). Reasoning with base rates is routine, relatively effortless, and context dependent. *Psychonomic Bulletin & Review*, *19*(3), 528–534. https://doi.org/10.3758/s13423-012-0249-3

Primi, C., Donati, M. A., Chiesi, F., & Morsanyi, K. (2018). Are there gender differences in cognitive reflection? Invariance and differences related to mathematics. *Thinking & Reasoning*, *24*(2), 258–279. https://doi.org/10.1080/13546783.2017.1387606

Raoelison, M., Thompson, V. A., & De Neys, W. (2020). The smart intuitor: Cognitive capacity predicts intuitive rather than deliberate thinking. *Cognition*, *204*, 104381. https://doi.org/10.1016/j.cognition.2020.104381

Raoelison, M., & De Neys, W. (2019). Do we de-bias ourselves?: The impact of repeated presentation on the bat-and-ball problem. *Judgment and Decision Making* *14*(2). https://doi.org/10.17605/OSF.IO/6AEC3

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. article. https://doi.org/10.1037/0033-295X.84.2.127

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, *6*, 532https://doi.org/10.3389/fpsyg.2015.00532

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Stanovich, K. E. (2018). Miserliness in human cognition: the interaction of detection, override and mindware. *Thinking & Reasoning*, *24*(4), 423–444. https://doi.org/10.1080/13546783.2018.1459314

Thompson, V. A. (2009). Dual process theory: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds; dual processes and beyond* (10th ed., pp. 171–195). Oxford University Press.

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, *20*(2), 215–244. https://doi.org/10.1080/13546783.2013.869763

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. S. B. T. (2018). Do smart people have better intuitions? *J Exp Psychol Gen*, *147*(7), 945–961. https://doi.org/10.1037/xge0000457

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001

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. https://doi.org/10.1016/j.cognition.2012.09.012

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Trémolière, B., De Neys, W., & Bonnefon, J.-F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*(3), 379–384. https://doi.org/10.1016/j.cognition.2012.05.011

Trippas, D., & Handley, S. J. (2017). The parallel processing model of belief bias: Review and extensions. In W. De Neys (Ed.), *Dual process theory 2.0.* (1st ed.). Routledge. https://doi.org/10.4324/9781315204550

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/bf03192720

Van Lier, J., Revlin, R., & De Neys, W. (2013). Detecting cheaters without thinking: Testing the automaticity of the cheater detection module. *PLoS One*, *8*(1), e53827https://doi.org/10.1371/journal.pone.0053827

Verschueren, N., Schaeken, W., & d'Ydewall, G. (2004). Everyday conditional reasoning with working memory preload. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26), 1399–1404. https://cloudfront.escholarship.org/dist/prd/content/qt7kk1x3qx/qt7kk1x3qx.pdf

Welsh, M. B., Burns, N. R., & Delfabbro, P. H. (2013). The cognitive reflection test: How much more than numerical ability? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35), 1587–1592. https://cloudfront.escholarship.org/dist/prd/content/qt68n012fh/qt68n012fh.pdf