Working Memory, Cognitive Miserliness and Logic as Predictors of Performance on the Cognitive Reflection Test

Edward J. N. Stupple (e.j.n.stupple@derby.ac.uk)
Centre for Psychological Research, University of Derby
Kedleston Road, Derby. DE22 1GB

Maggie Gale (m.gale@derby.ac.uk)
Centre for Psychological Research, University of Derby
Kedleston Road, Derby. DE22 1GB

Christopher R. Richmond (c.richmond1@unimail.derby.ac.uk)
Centre for Psychological Research, University of Derby
Kedleston Road, Derby. DE22 1GB

Abstract

The Cognitive Reflection Test (CRT) was devised to measure the inhibition of heuristic responses to favour analytic ones. Toplak, West and Stanovich (2011) demonstrated that the CRT was a powerful predictor of heuristics and biases task performance - proposing it as a metric of the cognitive miserliness central to dual process theories of thinking. This thesis was examined using reasoning response-times, normative responses from two reasoning tasks and working memory capacity (WMC) to predict individual differences in performance on the CRT. These data offered limited support for the view of miserliness as the primary factor in the CRT. The strongest predictor of CRT in both experiments was WMC. It is argued that while cognitive miserliness has been implicated in CRT performance, participants must also possess the requisite WMC and mindware to successfully complete it. Therefore, the psychological and psychometric properties of the CRT require continued study.

Keywords: Cognitive Reflection Test, Heuristics and Biases, Dual-process Theory, Belief-bias, Matching-bias, Reasoning, Cognitive Misers.

Introduction

Dual-process theories of reasoning and judgment dissociate fast and frugal 'snap' judgments from slow, effortful and methodical analyses (e.g., De Neys, 2012; Evans, 2007; Stanovich, 2004) with the latter being viewed as being more likely to lead to normatively sanctioned answers in a variety of reasoning tasks. These contrasting processes are captured by heuristic-analytic tasks that involve a conflict between these processes (see Kahneman, 2011 for a recent review) and are referred to as Type 1 (heuristic) and Type 2 (analytic) (e.g., Evans, 2011).

Frederick (2005) devised the Cognitive Reflection Test (CRT) to examine the ability of participants to resist intuitive, tempting answers in favour of deeper, more analytic ones. By way of illustration, an example item from the CRT is "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" Most participants respond that the answer is 10 cents; however, a slower and more analytic approach to the problem reveals the correct answer to be 5 cents.

The CRT has been a spectacular success, attracting more than 100 citations in 2012 alone (Scopus). This may be in part due to the ease of administration; with only three items and no requirement for expensive equipment, the practical advantages are considerable. There have, moreover, been numerous correlates of the CRT demonstrated, from a wide range of tasks in the heuristics and biases literature (Toplak et al., 2011) to risk aversion and SAT scores (Frederick, 2005). Its publication was also timely as it coincided with the recent boom in dual process theories of thinking and reasoning (e.g., De Neys, 2012; Evans, 2007; Stanovich, 1999). The CRT and its items have been adopted as a test-bed for the predictions of these theories (Bourgeois-Gironde, & Vanderhenst, 2009; Campitelli & Labollita, 2010; De Neys, Rossi, & Houdé, 2013; Toplak et al., 2011). Bourgeois-Gironde and Vanderhenst (2009) have also highlighted the advantage that the CRT offers in terms of testing dual process predictions against arithmetic norms rather than the more controversial normative standards in logic or probability (see Elqayam & Evans, 2011).

Toplak et al. (2011) presented perhaps the most comprehensive examination of the CRT, demonstrating considerable evidence for it as a predictor of non-normative responses to a battery of heuristics and biases tasks (each explicable) by dual process theories. Based on their findings Toplak et al. argued that the CRT predicts variance in rational thinking independently of intelligence, executive function and thinking dispositions, and that this variance is not insubstantial. Furthermore, Toplak et al. advance the CRT as a promising metric to tap into "What Intelligence Tests Miss" (Stanovich, 2009a) by accounting for rational thinking tendencies that are not captured by standard IQ tests (Stanovich suggests Dysrationalia as a term for people with higher IQ scores who fail on heuristics and biases tasks because they lack these thinking tendencies).

Stanovich (2009b) describes these rational thinking tendencies in a rational thinking taxonomy. Important categories for the CRT include, cognitive miserliness - the well-documented tendency to expend as little cognitive effort as is necessary to complete a task (first coined by Fiske & Taylor, 1984); and, 'mindware gaps' - whereby the necessary cognitive rules, strategies, or belief systems are lacking, corrupted or are not applied.

Moreover, De Neys, Rossi, and Houdé (2013) presented evidence in support of cognitive miserliness as an explanation of performance on the CRT, based on confidence ratings that demonstrated diminished confidence ratings for participants who give the '10 cents' response to the 'Bat and Ball' question. De Neys et al. argue that even though the participants had an intuitive sense of the correct response they still responded incorrectly. They explicitly argue that their data indicate that, while they appear to be cognitive misers, participants are not offering erroneous responses in blissful ignorance.

In further support of this position, Campitelli and Labollita (2010) investigated how individual differences in cognitive reflection impacted on decision-making. They argue that cognitive reflection is indicative of a thinking disposition related to Baron's (1988) proposals about Actively Open Minded Thinking. This thinking tendency is an obvious contrast with cognitive miserliness. Active Open Minded Thinking is associated with enhanced performance on a range of heuristics and biases tasks including the generation of alternatives and belief based reasoning tasks (Stanovich & West, 1999).

It would appear that the case for the CRT as a measure of cognitive miserliness is compelling. However, Thompson et al., (in press), examined the CRT as part of a paper testing the influence of perceptual fluency (Alter, Oppenheimer, Epley & Eyre, 2007) and answer fluency in priming deliberative thinking (Thompson, Prowse Turner, & Pennycook, 2011). They demonstrated that a degraded presentation of the CRT slowed participants down (conducive to analytic thinking and the converse of cognitive miserliness), but that this failed to facilitate correct responses among all but the most cognitively able participants (those in the uppermost quartile for IQ). These data suggest that increased response times to the CRT - which potentially ameliorate cognitive miserliness by encouraging greater cognitive effort - are not universally beneficial. These data, moreover, suggest that there is an important role for cognitive capacity (or working memory) in gaining the benefits of slower Type 2 processing.

In studies of syllogistic reasoning, response-times are predictive of normative responding, but this is not universal across problem types (Stupple, Ball, Evans and Kamal-Smith, 2011). Stupple, et al. demonstrated that inflated response times predicted normative responding where there were conflicts of belief and logic, and that this effect on normative responding was particularly associated with response times for invalid-believable problem types.

Further support for the utility of response times as a predictor of normative responding in tasks with a dual process conflict was reported by Stupple Ball and Ellis (2013), who created a heuristic-analytic conflict using matched and non-matched surface features in syllogistic reasoning problems (Stupple & Waterhouse, 2009). Stupple et al. (2013) noted that increased response times for invalid matching problems in a syllogistic reasoning task were associated with an increase in the overall normative responding. In contrast, increased response-times for valid non-matching problems were associated with *decreased* normative responding. These data demonstrate that it is not just the avoidance of miserliness that is important, but also that being sensitive to normative responses, perhaps by possessing the required mindware is important[1]. In short, a successful use of cognitive resources requires possession of the right mindware or the application of a sound strategy to be successful. Increased time deriving a response may indicate that Type 2 processing has occurred, but it is a fallacy to assume that the correct or normative answer will follow. A slow, effortful, but erroneous process cannot be characterized as the response of a miserly participant.

It is, nonetheless, argued that response times are vital to unpacking the predictions of dual process theories and that willingness to engage in time-consuming Type 2 processing on a syllogistic reasoning task should be predictive of willingness to engage in such processing on the CRT. A disposition to devote cognitive resources to a task coupled with the right mindware, however, may not be enough to find the correct answer if a participant has insufficient cognitive resource to reach the correct or normatively sanctioned conclusion.

Working memory capacity (WMC) has been shown to be important to reasoning performance, and to the process of analytic thinking (Bacon, Handley, Dennis & Newstead, 2008; Copeland, & Radvansky, 2004). Working Memory is, moreover, central to measures of intelligence (e.g., Kyllonen & Christal, 1990). Frederick (2005) makes sound arguments to differentiate the CRT from intelligence measures, but there is yet to be a detailed examination of the importance of WMC in solving the CRT. Detecting the error in the heuristic response to the CRT is arguably only the first step towards solving the problems in the CRT. Working out the correct response is likely to involve working memory demand, for example, when participants consider the candidate values for the ball and then concurrently calculate the total value of the bat and the ball. This argument is supported by the finding from Thompson et al. (in press) that Type 2 processing may only benefit the most cognitively able (and by implication the highest WMC) participants on the CRT. Toplak et al. (2011) argue that the items on the CRT are not insight problems (see Gilhooly & Murphy, 2005) – which do not incur significant working memory load – and are instead analytic problems, which do.

[1] Awareness of a 'double-negation elimination' logical rule (Rips, 1994) was proposed as important for reaching the normative answer for the problems used by Stupple et al. (2013).

Toplak et al. (2011) also acknowledge the influence of WMC and examine the role of CRT in predicting performance on heuristics and biases tasks, with the influence of WMC factored out. The focus here is instead upon the extent that WMC is predictive of the CRT in conjunction with Cognitive Miserliness, and sensitivity to normative considerations.

Mean response-times to syllogistic reasoning problems were used as an index of cognitive miserliness, a logic index (e.g., Stupple et al., 2013) was calculated to generate a measure of normative responding and a composite working memory score derived from Operation Span, Symmetry Span and Reading Span measures (Unsworth et al., 2005) was used as a measure of working memory capacity.

It is argued that Toplak et al.'s (2011) miserliness account of the CRT predicts that participants who devote the longest times to solving syllogisms would also be those who were most successful in solving CRT items. It was predicted that this would be the strongest predictor of CRT performance and the first factor included in the regression model by the stepwise procedure. It was also predicted that normative sensitivities and WMC would be significant predictors of CRT performance, but that these would account for less variance in CRT performance than the miserliness measure.

Method Experiment 1

Design Predictor variables were generated from the working memory span tasks (Unsworth et al., 2005) and the belief-bias reasoning task. Mean response-times to belief bias problems were calculated to generate an index of miserliness; acceptance rates for belief bias problems were used to generate a logic index. The dependent variable was the score on the CRT.

Participants Sixty-five undergraduates from the University of Derby, aged 18-45, were recruited via opportunity sampling. Participants had no training in formal logic and had not previously studied the psychology of reasoning or encountered the CRT. Each received a voucher (value £5) for participating.

Materials and procedure Participants received 16 target syllogisms counterbalanced for figure and mood. Belief-oriented contents were those employed by Stupple and Ball (2008). There were equal numbers of valid and invalid problems, and believable and unbelievable conclusions.

Logic index was calculated by adding acceptance rates for Valid Believable and Valid Unbelievable problems and subtracting total acceptance rate for Invalid Believable and Invalid Unbelievable problems (Valid Believable + Valid Unbelievable - Invalid Believable - Invalid Unbelievable).

Syllogisms and instructions were presented with Authorware 6.5 on a PC. Problems were counterbalanced, with contents rotated through them. WMC was measured using three complex span tasks (Unsworth et al., 2005) in E-Prime Version 2.0. These consisted of Automated Operation Span, Automated Symmetry Span and Automated Reading Span (see Unsworth et al., 2005 for details). The three measures of working memory capacity were combined to form a composite working memory score (Bartlett, 1937), derived from the three absolute span scores (defined as the sum of all sets of items that are recalled without error, Unsworth et al, 2005). The CRT was a pen and paper task.

Results Experiment 1

A Stepwise Multiple Regression tested the relative predictive strength of response-times and logic index in a belief-bias reasoning task and WMC for performance on the CRT. The Mean CRT score for the sample in Experiment 2 was 1.32 (SD= 1.11) which is well within the range described by Frederick (2005)[2].

Data indicated that WMC reliably accounted for 27% of the variability in CRT scores with participants with higher WMC scores performing better on the CRT than those with lower scores. Surprisingly, no further steps in the regression analysis significantly increased the variance accounted for as neither the Logic index nor the Response times were reliable predictors. Response-times demonstrated a non-significant correlation with CRT scores close to zero.

Table 1: Stepwise Regression Analysis of Working Memory Capacity, Logic index and Reasoning Response-times for Belief Bias problems as predictors of the CRT

Predictors	
Model 1	R^2= .28, R^2_{adj}=.27 F(1, 64)= 24.87, p<.001
WMC	β= .529, p< .001
Excluded	
Logic index	β= .156, p= .146
Response-times	β= .052, p= .641

Durbin Watson= 1.72, VIFs ranged from 1.01 to 1.08

Interim Discussion

These findings were contrary to the expectation as there was no reliable relationship shown for response times to syllogistic reasoning problems. Moreover, the variance explained by the composite measure of WMC was by far the most substantial predictor.

These results were surprising and may be specific to the syllogistic reasoning task employed. While there are similarities between belief bias problems and the CRT, in that some items may require the inhibition of an initial heuristic response, it is not the case that the CRT involves belief inhibition per se. A second experiment utilizing the same methodology, but employing an alternative set of reasoning problems (Roberts, 2005) that are also known to induce a heuristic-analytic conflict – the matching bias problems used by Stupple et al. (2013), was conducted.

[2] The average reported by Frederick was 1.24, N=3428.

Consistent with the CRT these matching problems feature conclusions, which are tempting to endorse, or reject based on their surface features. For example, in the case of the second item in the CRT: "If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes?" The most frequent erroneous response is 100 minutes, whereby participants may be matching their answer to the surface features of the problem. Similarly, performance on matching-bias syllogisms requires the inhibition of an inclination to respond based on whether surface features of conclusions and premises match, (and possessing the mindware to eliminate a double negation). It was hypothesized that (1) working memory capacity would again be a significant predictor of CRT scores, and, (2) that logic index and response times would predict CRT scores. However, these predictions were made with reduced confidence in the light of the findings from Experiment 1.

Method Experiment 2

Design Response times and conclusion acceptance rates from the matching bias reasoning task were used as predictors and the three Working Memory Span measures (Operation span, Reading span and Symmetry span (Unsworth et al., 2005) were again used to derive a composite WMC score. The dependent variable was the CRT scores.

Participants Forty-nine undergraduates from the University of Derby aged 18-45 were recruited via opportunity sampling. Participants had no training in formal logic and had not previously studied the psychology of reasoning or encountered the CRT. Each received a voucher (value £5) for participation.

Materials and Procedure Sixteen one-model syllogisms were presented. Conclusions either matched the premises (premises and conclusions were traditional affirmative or both were double negated), or were not matched with the premises - traditional affirmative premises were presented with double negated conclusions or double negated premises were presented with traditional affirmative conclusions. For non-conflict problems, analytic and heuristic strategies produced the same response, whereas for conflict problems analytic and heuristic matching strategies were in competition. Syllogism content involved combinations of professions and pastimes. These were rotated through the different problems. Reasoning problems, WMC measures and the CRT were administered identically to Experiment 1.

Results Experiment 2

The Mean CRT score for the sample in Experiment 2 was 1.12 (SD= 1.14) which is well within the range described by Frederick (2005), although for this experiment it was below the overall average reported by Frederick (2005).

A Stepwise Multiple Regression was conducted to test the relative predictive strength of response-times, and logic in a matching bias reasoning task and WMC for performance on the CRT. Data indicated that WMC reliably accounted for 23% of the variability in CRT scores in the first model, with participants with higher composite WMC scores demonstrating better performance on the CRT than those with lower scores. In a second model, the variance explained increased to 34% with the addition of the Logic index predictor. As with the first experiment, response-times did not reliably account for variance in CRT scores. More surprising, was the fact that the response-times correlated negatively (albeit unreliably) with CRT scores - in the opposite direction to that predicted.

Table 2: Stepwise Regression Analysis of Working Memory Capacity, Logic index and Reasoning Response-times for Matching Bias problems as predictors of the CRT

Predictors	
Model 1	R^2=.245, R^2_{adj}=.229 F(1, 48)= 15.54, p=.001,
WMC	β=.495, p<.001
Model 2	R^2=.365, R^2_{adj}=.338 F(2, 47)=13.52, p=.001, F change, p=.004
WMC	β=.426, p=.001
Logic index	β=.354, p=.004
Excluded	
Response-times	β=-.149, p=.203

Durbin Watson= 1.70, VIF = 1.04

Discussion

The experiments presented here tested the relative contributions of Response times to reasoning tasks (as an index of cognitive miserliness), Logic Index (as a measure of sensitivity to normative responses) and WMC to predicting variance in the CRT. Consistent with predictions, WMC was a reliable predictor of performance on the CRT in both experiments - and was a substantially stronger predictor than expected. Moreover, the unexpected null finding for response times, suggested that if the CRT is conceptualized as a measure of cognitive miserliness then it might not convincingly generalize beyond the arithmetic based problems to standard dual processing tasks such as belief bias or matching bias syllogisms. If the CRT is a general measure of cognitive miserliness then those participants responding primarily with the Type 2 answers to the CRT should engage in more Type 2 processing on syllogistic reasoning tasks as indexed by increased response times. These data suggest that this was not reliably the case.

WMC correlating most strongly with performance on the CRT is somewhat problematic for the use of the test as a measure of miserliness. Individuals with lower WMC may expend a great deal of effort in attempting to solve heuristic-analytic problems, but lack the capacity to maintain their representation of, for example, possible ball costs relative to the bat as they work through the alternatives. Participants with higher WMC may find the cognitive costs less expensive and, thus be more willing to pay them[3].

Cognitive miserliness could be argued to be relative to the cognitive resource of the participant. A participant with a high WMC who provides heuristic responses to the CRT would be categorized appropriately as a cognitive miser as they had the necessary cognitive resources, but chose not to apply them to the task. In contrast, a participant with lower WMC who devotes considerable time and effort, but arrives at a heuristic answer would be inappropriately described as miserly (perhaps they could be considered cognitive wastrels instead). It may be that those participants with greater WMC can engage in the deliberative thought required to avoid the heuristic response with relatively less effort when compared to those with lesser WMC. This reduced cognitive cost may become affordable to the participants with more miserly tendencies, but who also have ample working memory resources available[4].

De Neys et al. (2013) suggest that participants are aware of the incongruity of answering 10 cents to the bat and ball question, but often fail to engage the deliberative processing required for the correct ‘5 cents’ answer. We would add to this claim that while cognitive miserliness is almost certainly a factor, our data indicate that, for a proportion of participants at least, they may not have the cognitive resources to pursue their metacognitive uncertainty about their intuitive response. Alternatively, the intuitive response may offer a cognitive escape hatch, if processing demands are too great (cf. Quayle & Ball, 2000). Similarly, with regard to Thompson et al.'s (in press) findings - that only the most able participants benefitted from the Type 2 processing that dis-fluent stimuli encouraged in terms of the accuracy of their responding demonstrating that increased response times may be important to success on the CRT, but they are not sufficient. Further investigation is required to understand the nuanced interplay between miserliness and cognitive ability/working memory capacity on the CRT.

It was notable that there was not a reliable relationship between normative responding in a belief bias task with performance on the CRT, but that normative responding on the matching bias task was a highly reliable predictor of CRT performance. A possible account of the discrepancy between studies could be based on the manner in which the heuristic-analytic conflict is resolved. Optimal performance on belief bias problems requires an ability to inhibit belief driven responses while searching for alternative models (Stupple et al., 2011), whereas the matching bias problems required an explicit awareness of the logic of double negatives - such that errors could be characterized as the result of missing or corrupted 'mindware' (Stanovich, 2009b). This difference in the source of the heuristic-analytic conflict could potentially account for the discrepancy between problem types. This further contrasts with the proposal of the CRT as an index of cognitive miserliness. The absence of the appropriate mindware for double negations among those participants who score lowest on the CRT would appear to indicate a lack of an understanding of logic or rule based thinking, rather than, an unwillingness to engage in the requisite cognitive effort. This is inconsistent with the arguments from Toplak et al. (2011), who suggest that knowledge gaps represent a major class of reasoning error but that: "The potency of the CRT as a predictor of performance on heuristics-and-biases tasks certainly does not derive from its ability to assess knowledge gaps, because it clearly does no such thing." (Toplak et al, 2011, p. 1284). The variance in CRT scores explained by normative responding to matching-bias syllogisms cannot reasonably be claimed as a causal link, but suggests an association between possessing the necessary cognitive rules or strategies for detecting matching bias conflicts and the heuristic-analytic conflicts that are implicated in success on the CRT. We would argue that examination of the CRT as an index of conflict detection also warrants further investigation.

Therefore, it is advocated that self-report measures such as the Need for Cognition (Cacioppo & Petty, 1982) or Rational-Experiential Inventory (Epstein, 1994) continue to be used alongside the CRT to quantify the subjective experience of miserliness. This subjective experience is likely to co-vary with cognitive capacity - relative to the task demands. The CRT as an index of cognitive miserliness presupposes a degree of equality in our cognitive wealth. Self-report measures may supplement the CRT by offering insight into the experience of how effortful the task was and by quantifying self-perceptions of cognitive miserliness.

Nonetheless, we agree with Toplak et al. (2011) that the CRT captures variability in performance on heuristics and biases tasks that are not captured by IQ tests and the CRT remains a promising measure to explore in this regard. What remains clear from these data is that explaining the psychological properties of the CRT is not a simple task, and while it is undoubtedly an influential task that will remain popular among dual process theorists, the precise nature of its psychometric and psychological properties require continued study.

Acknowledgments

This research was funded through the Research Inspired Curriculum Fund at the University of Derby.

[3] A reviewer suggested that high WMC individuals might solve problems more rapidly and thus not show the anticipated correlation. However, when WMC is controlled for there is still no reliable correlation between CRT scores and reasoning task response times (Exp. 1, p=.64; Exp. 2, p=.22). However, this possibility warrants a fine-grained examination in future.

[4] It was also suggested - based on Kuhl (2000) – that some ‘high logic’ participants prematurely inhibit alternative construals of CRT questions to avoid ambiguity, and this explains some variance in CRT scores – again, this warrants further investigation.

References

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, 569–576.

Bacon, A.M., Handley, S.J., Dennis, I. & Newstead, S.E. (2008) Reasoning strategies: the role of working memory and verbal-spatial ability. *European Journal of Cognitive Psychology* 20 (6), 1065 – 1086

Baron, J. (1988). *Thinking and deciding.* New York: Cambridge University Press. (4th edition).

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology, 28,* 97-104.

Bourgeois-Gironde, S., & Vanderhenst, J. B. (2009). How to open the door to System 2: Debiasing the Bat and Ball problem. In S. Watanabe, A.P Bloisdell, L. Huber, A. Young (Eds.), *Rational animals, irrational humans*. Tokyo: Keio University Press.

Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116-131.

Campitelli, G., & Labollita, M. (2010) Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5, 182–191.

Copeland, D. & Radvansky, G. (2004). Working memory and syllogistic reasoning, *The Quarterly Journal of Experimental Psychology* Section A, 57:8, 1437-1457

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7 (1), 28-38.

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools**.** *Psychonomic Bulletin & Review*

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral & Brain Sciences*, *34*, 233-248.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709-724

Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13,* 321-339.

Evans, J. St. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, 31, 2, 86–102

Fiske, S. T., Taylor, S. E. (1984). *Social cognition*. MA: Addison-Wesley Pub. Co

Gilhooly, K. .J., & Murphy, P.(2005). Differentiating insight from non-insight problems. *Thinking & Reasoning* 11 (3), 279-302.

Kahneman, D. (2011) *Thinking, Fast and Slow*. London: Allen Lane.

Kuhl, J. (2000). A functional-design approach to motivation and self-regulation: The Dynamics of personality systems and interactions. In M. Boekaerts & P.R. Pinrich (Eds.), *Handbook of Self-regulation*. San Diego, Academic Press.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *53A*, 1202-1223.

Rips, L. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Roberts, M. J. (2005). Expanding the universe of categorical syllogisms: A challenge for reasoning researchers. *Behavior Research Methods*, *37*, 560-580.

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

Stanovich, K. E. (2009a). *What intelligence tests miss: The psychology of rational thought.* New Haven: Yale University Press.

Stanovich, K. E. (2009b). Distinguishing the reflective, algorithmic and autonomous minds: Is it time for a tri-process theory? In J.St.B.T.Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, *14*, 168-181.

Stupple, E .J. N., Ball, L. J., Evans, J. St. B. T., & Kamal-Smith, E. N. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology*, *23*, 931-941.

Stupple, E.J.N. Ball, L.J. & Ellis, D. (2013). Matching bias in syllogistic reasoning: Evidence for a dual-process account from response times and confidence ratings. *Thinking & Reasoning*, 19(1):54-77

Stupple, E. J. N., & Waterhouse E. F. (2009). Negations in syllogistic reasoning: evidence for a heuristic-analytic conflict. *Quarterly Journal of Experimental Psychology*, *62*, 1533-1541.

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

Thompson, V. A., Prowse Turner, J. A., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (In press) The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*

Toplak, West and Stanovich, K.E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39, 7, 1275-1289

Unsworth, N., Heitz, R.P., Schrock, J. C., & Engle, R. W. (2005) An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505.