

The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks

Maggie E. Toplak · Richard F. West · Keith E. Stanovich

Published online: 4 May 2011
© Psychonomic Society, Inc. 2011

Abstract The Cognitive Reflection Test (CRT; Frederick, 2005) is designed to measure the tendency to override a prepotent response alternative that is incorrect and to engage in further reflection that leads to the correct response. In this study, we showed that the CRT is a more potent predictor of performance on a wide sample of tasks from the heuristics-and-biases literature than measures of cognitive ability, thinking dispositions, and executive functioning. Although the CRT has a substantial correlation with cognitive ability, a series of regression analyses indicated that the CRT was a unique predictor of performance on heuristics-and-biases tasks. It accounted for substantial additional variance after the other measures of individual differences had been statistically controlled. We conjecture that this is because neither intelligence tests nor measures of executive functioning assess the tendency toward miserly processing in the way that the CRT does. We argue that the CRT is a particularly potent measure of the tendency toward miserly processing because it is a performance measure rather than a self-report measure.

Keywords Cognitive reflection test · Rational thinking · Intelligence · Heuristics and biases · Thinking dispositions

M. E. Toplak (✉)
Department of Psychology, York University,
4700 Keele Street, 126 BSB,
Toronto, Ontario, Canada M3J 1P3
e-mail: mtoplak@yorku.ca

R. F. West
Department of Psychology, James Madison University,
Harrisonburg, VA, USA

K. E. Stanovich
Department of Human Development and Applied Psychology,
University of Toronto,
Toronto, Ontario, Canada

The Cognitive Reflection Test (CRT) is a three-item measure introduced into the journal literature by Frederick (2005). The task is designed to measure the tendency to override a prepotent response alternative that is incorrect and to engage in further reflection that leads to the correct response. The quintessential item from the CRT was first discussed by Kahneman and Frederick (2002) in an article that reframed the heuristics-and-biases literature in terms of the concept of attribute substitution. The problem is as follows: A bat and a ball cost \$1.10 in total. The bat costs \$1 more than the ball. How much does the ball cost?

When they answer this problem, many people show a characteristic that is common to many reasoning errors: They behave like cognitive misers (Dawes, 1976; Simon, 1955, 1956; Stanovich, 2009b; Taylor, 1981; Tversky & Kahneman, 1974). They give the first response that comes to mind—10 cents—without thinking further and realizing that this cannot be right. The bat would then have to cost \$1.10, and the total cost would then be \$1.20 rather than the required \$1.10. People often do not think deeply enough to realize their error, and cognitive ability is no guarantee against making the error. Frederick (2005) found that large numbers of highly select university students at MIT, Princeton, and Harvard were cognitive misers; they responded that the cost was 10 cents, rather than the correct answer... 5 cents.

This problem and the two others (see the [Method](#) section below) on the CRT seem at first glance to be similar to the well-known insight problems in the problem-solving literature, but they in fact display a critical difference. Classic insight problems (see Gilhooly & Fioratou, 2009; Gilhooly & Murphy, 2005) do not usually trigger an attractive alternative response. Instead, the participant sits lost in thought trying to reframe the problem correctly—in, for example, the classic nine-dot problem. The three

problems on the CRT are of interest to researchers working in the heuristics-and-biases tradition because a strong alternative response is initially primed and then must be overridden. As Kahneman and Frederick made clear in their 2002 paper, this framework of an incorrectly primed initial response that must be overridden fits in nicely with currently popular dual-process frameworks (De Neys & Glumicic, 2008; Evans, 1984, 2008, 2010; Evans & Frankish, 2009; Lieberman, 2007, 2009; Sloman, 1996, 2002; Stanovich, 1999, 2009a, 2011). Kahneman (2000) pointed out that such a framework had been an underlying assumption of his earlier work with Tversky.

The CRT would seem to be ideally constructed as a predictor of performance on heuristics-and-biases tasks, but the data have been inconsistent. Frederick (2005) observed that with as few as three items, his CRT could predict performance on measures of temporal discounting, the tendency to choose high-expected-value gambles, and framing effects. Likewise, Cokely and Kelley (2009) found a correlation of .27 between performance on the CRT and the proportion of choices consistent with expected value. In contrast, Campitelli and Labollita (2010) found little relation between CRT performance and the choice of high-expected-value gambles. Oechssler, Roider, and Schmitz (2009) found the CRT to be related to the number of expected-value choices and the tendency to commit the conjunction fallacy. In contrast, Obrecht, Chapman, and Gelman (2009) found no relation between CRT performance and the degree of encounter frequency bias. Finally, Koehler and James (2010) found significant correlations between the CRT and the use of and endorsement of maximizing strategies on probabilistic prediction tasks.

In the present article, we explore the predictive properties of the CRT in a much wider range of the heuristics-and-biases tasks. Additionally, however, we attempt to uncover some of the underlying psychological structure of the CRT. This is necessary because on the surface, the CRT appears to be a somewhat complex measure. It seems to carry properties across the boundary of an important distinction in classical personality and psychometric work—that is, the distinction between cognitive abilities and thinking dispositions. This conceptual distinction follows from differentiating optimal (sometimes termed *maximal*) performance situations and typical performance situations (see Ackerman, 1994, 1996; Ackerman & Heggstad, 1997; Ackerman & Kanfer, 2004; see also Cronbach, 1949; Matthews, Zeidner, & Roberts, 2002). Typical performance situations are unconstrained, in that no overt instructions to maximize performance are given, and the task interpretation is determined to some extent by the participant. The goals to be pursued in the task are left somewhat open. The issue is what a person would typically do in such a situation, given few constraints (see Stanovich, 2009b). In contrast, optimal

performance situations are those in which the task interpretation is determined externally (not left to the participant). The person performing the task is instructed to maximize performance. Duckworth (2009) has discussed the surprisingly weak relation between typical and maximal performance across a variety of domains. For example, Sackett, Zedeck, and Fogli (1988) found that there were very low correlations between the maximal item-processing efficiency that supermarket cashiers could attain and the typical processing efficiency that they usually attained.

All tests of intelligence or cognitive aptitude are optimal performance assessments, whereas measures of thinking dispositions are often assessed under typical performance conditions (Ackerman & Heggstad, 1997; Cacioppo, Petty, Feinstein, & Jarvis, 1996; Norris & Ennis, 1989; Perkins, 1995; Sternberg, 2003; Zeidner & Matthews, 2000). The CRT, in fact, may derive its potency as a predictor from the fact that it taps both a cognitive ability dimension and a thinking disposition dimension. Frederick (2005) reported a correlation of .44 between CRT performance and SAT total scores, as well as a .43 correlation between CRT scores and performance on the Wonderlic IQ test. Obrecht, Chapman, and Gelman (2009) observed a correlation of .45 between performance on the CRT and SAT quantitative scores, and below we report a .40 correlation between cognitive ability and CRT performance. The CRT clearly has moderate overlap with measures of cognitive ability.

Despite these indications of correlations with cognitive ability measures, on a face validity basis, the CRT appears to also implicate thinking dispositions—particularly those related to reflectivity, the tendency to engage in fully disjunctive reasoning, and the tendency to seek alternative solutions. In the present study, we attempted to partition the predictive variance of the CRT by examining its ability to predict a wider range of heuristics-and-biases and judgment-and-decision-making tasks than has been investigated in previous research. We also examined its ability to predict the degree of belief bias in syllogistic reasoning. Our study investigated whether the variance that the CRT shares with these measures of rational thinking is also shared by cognitive ability and a selection of thinking dispositions. In addition, we also examined another class of variable that may help to reveal the underlying psychological structure of the CRT. Recent work on the inhibitory and set-shifting properties of executive-functioning tasks makes this class of processes a potentially theoretically interesting correlate of performance on the CRT (Aron, 2008; Best, Miller, & Jones, 2009; Duncan et al., 2008; Friedman et al., 2007; Hasher, Lustig, & Zacks, 2007; Miyake, Friedman, Emerson, & Witzki, 2000; Salthouse, Atkinson, & Berish, 2003; Zelazo, 2004). As the bat/ball example described above illustrates, answering the problems on the CRT requires suppressing a prepotent “natural” (see Kahneman,

2003) response to the problem. Such suppression could well be related to the types of set-shifting and inhibitory processes that are directly and indirectly assessed on measures of executive functioning. We thus included three executive-functioning tasks in our study to complement the cognitive ability measures and thinking dispositions that were used to examine the reasons that the CRT predicts performance on tasks used in the heuristics-and-biases literature. Our heuristics and biases tasks spanned the gamut of this vast literature, as we shall now describe in the [Method](#) section.

Method

Participants and procedure

A total of 346 participants (95 males and 251 females; mean age = 20.1 years, $SD = 3.9$) took part in the study. The majority of these students were first-year undergraduates (223 students); 52 of the students were in their second undergraduate year, 29 were in their third undergraduate year, 30 were in their fourth undergraduate year, and 12 had completed their undergraduate degree. The participants were recruited at a large university and were either part of a participant pool who received course credit or were paid for their participation. There were no age or gender differences between the paid and unpaid participants. Participants completed the battery of tasks described below, plus some other measures, during a single, 2-h session.

Tasks and variables

Cognitive reflection test

Taken from Frederick (2005), this test is composed of three questions, as follows:

- A bat and a ball cost \$1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ____ cents
- If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ min
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days

What characterizes these problems is that a quick, intuitive answer springs to mind, but that this quick answer is incorrect. The key to deriving the correct solution is to suppress and/or evaluate the first solution that springs to mind (Frederick, 2005). The solution to the bat-and-ball

problem is 5 cents, to the widget problem is 5 min, and to the lily pad problem is 47 days. Our problems were run without the prior instructions given by Frederick (2005): “Below are several problems that vary in difficulty. Try to answer as many as you can.” A composite measure of performance on these three items was used as the dependent measure. Mean performance was 0.7 items correct ($SD = 0.93$); 55.8% ($n = 193$) of participants did not solve any of the problems, and 6.6% ($n = 23$) solved all three items.

Cognitive ability

The Vocabulary and Matrix Reasoning subtests from the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) were used as indices of verbal and nonverbal ability. The mean raw score on the Vocabulary subtest was 52.6 ($SD = 7.4$), and the mean raw score of the Matrix Reasoning subtest was 27.3 ($SD = 3.7$). The raw scores for the Vocabulary and Matrix Reasoning subtests were converted into z scores and summed to create a composite measure of cognitive ability.

Heuristics-and-biases tasks

A group of 15 classic heuristics-and-biases tasks were chosen that reflected important aspects of rational thought, including probabilistic reasoning, hypothetical thought, theory justification, scientific reasoning, and the tendency to think statistically. The heuristics-and-biases battery consisted of one causal base-rate problem, two sample-size problems, one problem assessing sensitivity to regression to the mean, two gambler’s fallacy problems, one conjunction problem, one covariation detection problem, one methodological reasoning problem, one Bayesian reasoning problem, a framing problem, one problem assessing denominator neglect, a probability matching assessment, a sunk cost problem, and an outcome bias problem. A description of each of the problems is presented in the [Appendix](#).

Each of the 15 problems in the heuristics-and-biases battery was scored 0 or 1 (see the [Appendix](#) for a description of the scoring of each item), and the scores were summed to form a composite score ($M = 6.88$, $SD = 2.32$). By forming a composite score, we do not mean to imply that these heuristics-and-biases tasks form a strong unidimensional construct. The rational-thinking tendencies measured by these heuristics-and-biases tasks are probably multifarious (Reyna, Lloyd, & Brainerd, 2003; Stanovich, 2009b, 2011; Stanovich, West, & Toplak, 2011). Nevertheless, previous research has indicated some degree of common variance among them (Bruine de Bruin, Parker, & Fischhoff, 2007; Finucane & Gullion, 2010; Klaczynski, 2001; Parker & Fischhoff, 2005; Slugoski, Shields, & Dawson, 1993;

Stanovich & West, 1998c, 2000; West, Toplak, & Stanovich, 2008). However, each task, from a psychometric point of view, represents only a single item. Of the 105 possible correlations among the heuristics-and-biases tasks, 86 were in the positive direction, but only 39 significantly so. Thus, only modest reliability for the composite score was expected, and this was the case. The split-half reliability was .495, and Cronbach's alpha was .484.

Syllogistic reasoning problems with belief bias

Two sets of syllogistic reasoning tasks were presented in different parts of the reasoning battery. The first set included three deductive reasoning items in which the believability of the conclusion was pitted against the validity of the argument (Evans, Barston, & Pollard, 1983; Sá, West, & Stanovich, 1999). One of the items had the following structure: "All living things need water. Roses need water. Conclusion: Roses are living things." Participants were asked to determine whether the conclusion did or did not follow from the premises. In each of the three problems, the believability of the conclusion was inconsistent with the validity of the argument. For example, in this sample item, the problem has a believable conclusion, but the argument is invalid.

Two other problems were used, which were based on the work of George (1995), who designed a deductive reasoning task that assesses whether participants recognize the deductive certainty of modus ponens. One example went as follows: "Premises: 1. If a car is a Honda, then it is expensive. 2. John's car is a Honda. Conclusion: 3. John's car is expensive." Participants responded on the following scale after reading instructions similar to those used on the previous three problems: true, probably true, somewhat true, uncertain, somewhat false, probably false, and false. Responding "true" was scored as 1, and any other response was scored as 0. Across the five reasoning problems, the mean number correct was 2.72 ($SD = 1.21$).

Executive-functioning measures

Set shifting The Trailmaking Test (Reitan, 1955, 1958) requires the participant to connect 13 numbered and 12 lettered circles. The participant is instructed to alternate between numeric and alphabetic order, going from 1 to A to 2 to B to 3 to C, and so forth. The mean completion time was 59.6 s ($SD = 24.6$ s). After a square-root transformation, the scores were transformed to z scores, and the z scores were reflected so that higher scores indicated better set-shifting ability.

Inhibition The Stroop task was used to measure inhibition. There were three different conditions, each with 24 items

arranged in a 4 x 6 matrix: a word-reading condition, a color-naming condition, and an interference condition. The dependent variable of the Stroop task was the total naming time (in seconds) for the interference condition minus the total naming time for the color condition. The mean interference score was 10.2 s ($SD = 5.0$, range 0.1 to 27.1). These scores were standardized, and the z scores were reflected so that higher scores indicated better ability to inhibit.

Working memory We used the Paced Auditory Serial Addition Test (PASAT; Gronwall, 1977) as our measure of working memory. It is a serial-addition task used to assess working memory, divided attention, and information-processing speed (Gonzalez et al., 2006; Strauss, Sherman, & Spreen, 2006). In this task, a computer is used to serially present single digits at a rate of one digit every 3 s (Trial 1) and every 2 s (Trial 2). A practice trial precedes each of the actual trials. In each trial, the participant must add each new digit to the one immediately prior to it. The dependent measure was the total number of correct sums given, out of a possible 60, during each trial. An average score was calculated for Trials 1 and 2, resulting in a mean performance of 38.3 ($SD = 9.3$). Standardized z scores were used as the dependent measure on this task.

Because working memory is often as or more highly correlated with cognitive ability measures than with executive-functioning measures, we created another cognitive ability index (CA2) with working memory as a component. The standard scores of the WASI composite and the working memory task were summed to form this second, CA2, composite index of cognitive ability.

Thinking dispositions

Participants completed a self-report questionnaire in which they were asked to rate their agreement with each question using the following 6-point scale: (1) *strongly disagree*, (2) *disagree moderately*, (3) *disagree slightly*, (4) *agree slightly*, (5) *agree moderately*, and (6) *strongly agree*. Questions were presented in mixed order.

The first thinking dispositions measure was the Actively Openminded Thinking scale (Stanovich & West, 1997, 2007), which is a 41-item measure scored so that higher scores represented a greater tendency toward open-minded thinking. Examples of items are "People should always take into consideration evidence that goes against their beliefs," "Certain beliefs are just too important to abandon, no matter how good a case can be made against them" (reverse scored), and "No one can talk me out of something I know is right" (reverse scored). The score on the scale was obtained by summing the responses to the 41 items ($M = 161.1$, $SD = 19.6$). The split-half reliability of

the scale (Spearman–Brown corrected) was .78, and Cronbach’s alpha was .81.

Superstitious thinking has been found to predict probabilistic reasoning (Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Toplak, Liu, Macpherson, Toneatto, & Stanovich, 2007). Our superstitious thinking scale was composed of two items from a paranormal scale used by Jones, Russell, and Nickel (1977), four items from a luck scale used by Stanovich and West (1998c), four items from an ESP scale used by Stanovich (1989), and three items from a superstitious thinking scale published by Epstein and Meier (1989). Examples of items included “Astrology can be useful in making personality judgments,” “The number 13 is unlucky,” and “I do not believe in any superstitions” (reverse scored). The score on the scale was obtained by summing the responses to the 13 items ($M = 33.5$, $SD = 10.4$). The split-half reliability of the scale (Spearman–Brown corrected) was .83, and Cronbach’s alpha was .81. Scores on the superstitious thinking scale were reflected so as to go in the same direction as the other two thinking disposition measures.

The Consideration of Future Consequences (CFC) scale is a 12-item scale that was developed by Strathman, Gleicher, Boninger, and Scott Edwards (1994) to measure the extent to which individuals consider distant outcomes when choosing their present behavior. A sample item from the scale was “I only act to satisfy immediate concerns, figuring the future will take care of itself” (reverse scored). The score on the scale was obtained by summing the responses to the 12 items ($M = 48.2$, $SD = 7.4$). The split-half reliability of the scale (Spearman–Brown corrected) was .53, and Cronbach’s alpha was .55.

Results

Table 1 displays the percentages of participants who responded correctly on each of the heuristics-and-biases tasks. There is considerable variation in task difficulty. The most difficult task was the sample-size squash problem, answered correctly by only 15.6% of the participants, and the easiest task was the second gambler’s fallacy problem, which was answered correctly by 92.2% of the participants. None of the 13 remaining tasks was answered correctly by more than 75% of the participants. This is significant because, collectively, these tasks assess whether people adhere to some of the most fundamental strictures of rational thought (see Baron, 2008; Bishop & Trout, 2005; Evans & Over, 1996; Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 1996, 2000; Samuels & Stich, 2004; Shafir & LeBoeuf, 2002; Stanovich, 1999, 2004, 2011).

These results converge with a body of other work indicating that the susceptibility to these biases varies considerably (Bruine de Bruin et al., 2007; Cokely &

Table 1 Percentages of correct responses on each of the heuristics-and-biases tasks

Task	% Correct Responses
1. Causal Base Rate	53.8
2. Sample Size: Hospital Problem	28.6
3. Sample Size: Squash Problem	15.6
4. Regression to the Mean	40.2
5. Gambler’s Fallacy 1	69.4
6. Gambler’s Fallacy 2	92.2
7. Conjunction Problem	19.1
8. Covariation Detection	36.1
9. Methodological Reasoning	27.2
10. Bayesian Reasoning	22.3
11. Framing Problem	62.7
12. Probabilistic Reasoning: Denominator Neglect	61.8
13. Probability Matching	23.5
14. Sunk Cost	64.2
15. Outcome Bias	72.0

Kelley, 2009; Del Missier, Mantyla, & Bruine de Bruine, 2010; Dohmen, Falk, Huffman, Marklein, & Sunde, 2009; Klaczynski, 2001; Oechssler et al., 2009; Stanovich & West, 1998a, 1998c, 1999, 2000, 2008b; West et al., 2008). What predicts this variation in susceptibility to different biases, and how does this variation relate to that in another foundational critical thinking skill—reasoning independently of prior belief (the syllogistic reasoning task)? The next several analyses address these questions in various ways.

Table 2 presents the zero-order correlations among the major variables in the study. Because of the large sample size in the study, all correlations over .125 are significant at the .01 level (one-tailed). The two components of rational thinking—avoidance of thinking biases on the heuristic-and-biases tasks and syllogistic reasoning independent of prior belief—displayed a moderate correlation with each other (.29). These two variables were standardized and added together to form a rational-thinking composite score. The CRT displayed its highest correlation with this composite score (.49), followed by its correlation with the heuristics-and-biases composite (.42) and its correlation with CA2 (.40), the cognitive ability indicator that combined the WASI composite with working memory performance. Thus, two characteristics of the CRT appear to be that it has moderate overlap with cognitive ability and that it is a predictor of rational thinking. We explore the correlates of the latter point next.

In terms of zero-order correlations, it is clear from Table 2 that the strongest correlate of performance on the rational-thinking composite score was, in fact, the CRT ($r = .49$). Cognitive ability was the next most potent zero-order predictor. The WASI displayed a correlation of .41 with

Table 2 Correlations between Cognitive Reflection Test, rational-thinking tasks, cognitive ability measures, executive-function measures, and thinking dispositions measures

Task	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Cognitive Reflection Test	–												
Rational-Thinking Measures													
2. Heuristic-and-Biases Composite	.42	–											
3. Belief Bias Syllogistic Reasoning	.36	.29	–										
4. Rational-Thinking Composite	.49	(.80)	(.80)	–									
Cognitive Ability Measures													
5. WASI Composite	.32	.32	.35	.41	–								
6. WASI Vocabulary	.17	.24	.32	.35	(.76)	–							
7. WASI Matrix Reasoning	.32	.24	.21	.28	(.76)	.16	–						
8. CA2 (WASI Comp+ WM)	.40	.39	.36	.47	(.81)	(.66)	(.71)	–					
Executive-Function Measures													
9. Set Shifting	.28	.15	.12	.17	.27	.11	.30	.40	–				
10. Inhibition	.15	.17	.16	.21	.24	.18	.19	.35	.32	–			
11. Working Memory	.33	.33	.22	.34	.30	.17	.29	(.80)	.41	.36	–		
Thinking Dispositions													
12. Actively Open-Minded Thinking	.10	.18	.12	.19	.21	.25	.07	.17	.07	.09	.02	–	
13. Superstitious Thinking	.04	.24	.04	.18	.12	.22	.04	.15	.08	.10	.14	.38	–
14. CFC	.10	.16	.15	.19	.26	.23	.16	.21	.06	.04	.03	.21	.12

Signs have been changed so that all correlations indicate values in expected direction. Correlations in parentheses reflect part-whole relationships. $r = .088$ is significant at $p < .05$; $r = .125$ is significant at $p < .01$; $r = .165$ is significant at $p < .001$, one-tailed

the rational-thinking composite score and the summed standard scores of the WASI and the working memory task (CA2) displayed a correlation of .47. The executive-function measures and the thinking dispositions measures displayed smaller but significant correlations (.17 to .34 and .18 to .19, respectively) with the rational-thinking composite score.

With few exceptions, the patterns of prediction were similar for the heuristics-and-biases tasks and the syllogistic reasoning task taken separately. The CRT was the strongest correlate of the former ($r = .42$) and was tied with CA2 ($r = .36$) as the most potent predictor of the latter. Most measures were more correlated with performance on the heuristics-and-biases tasks than with performance on syllogistic reasoning with belief bias. This was particularly true of the thinking dispositions measures (.16 to .24 versus .04 to .15).

The correlations displayed in Table 2 indicate that variance in CRT performance overlaps with both intelligence and rational-thinking ability. This finding of course raises the question of whether the CRT predicts rational thinking merely because of its association with cognitive ability. The next series of analyses explore whether, with respect to predicting rational-thinking ability, the predictive variance of the CRT is entirely redundant with that of cognitive ability. In short, these analyses assess whether the CRT measures properties relevant to rational thinking that go beyond those measured on intelligence tests or the other

factors examined here: executive-functioning measures and thinking dispositions.

The regression analyses in Table 3 explore how the predictive variance of the CRT overlaps with that of cognitive ability, executive-function measures, and thinking dispositions. The criterion variable in the first hierarchical regression analysis was the rational-thinking composite score. The first block of variables entered were the WASI Vocabulary and WASI Matrix scores, and they accounted for 17.3% of the variance ($p < .001$). The second block of variables entered were the three executive-functioning measures, and they accounted for an additional 5.6% of the variance ($p < .001$). Entered third as a block were the three thinking disposition measures, and they accounted for an additional 2.1% of the variance ($p < .05$). Finally, scores on the CRT were entered into the equation and accounted for a substantial amount of unique variance (11.2%, $p < .001$).

The results of this analysis clearly indicate that the CRT's ability to predict performance on rational-thinking tasks is not entirely due to its variance in common with cognitive ability—nor is it due to its variance in common with executive functioning in addition to cognitive ability. Finally, when the overlap with thinking disposition measures is partialled out as well, the CRT remains able to predict substantial unique variance. In the far right column of Table 3 is listed the unique variance accounted for by each of the blocks when

Table 3 Regression results

	Hierarchical		Unique Variance Explained
	R^2 Change	F to Enter	
Criterion Variable = Rational-Thinking Composite			
1. Intelligence Block: WASI Vocab; WASI Matrix	.173	35.98	.042
2. Executive-Function Block: Set-S; Inhibition; WM	.056	8.26	.025
3. Thinking Dispositions Block: AOT; Super; CFC	.021	3.12	.017
4. Cognitive Reflection Test	.112	59.01	.112
Overall Regression: $F = 21.22$			
Multiple $R = .602$			
Multiple $R^2 = .362$			
Criterion Variable = Heuristics-and-Biases Composite			
1. Intelligence Block: WASI Vocab; WASI Matrix	.100	19.09	.013
2. Executive-Function Block: Set-S; Inhibition; WM	.060	8.07	.026
3. Thinking Dispositions Block: AOT; Super; CFC	.042	5.96	.040
4. Cognitive Reflection Test	.080	37.45	.080
Overall Regression: $F = 14.68$			
Multiple $R = .531$			
Multiple $R^2 = .282$			
Criterion Variable = Belief Bias Syllogistic Reasoning			
1. Intelligence Block: WASI Vocab; WASI Matrix	.128	25.10	.051
2. Executive-Function Block: Set-S; Inhibition; WM			
Set-S; Inhibition; WM	.019	2.49 ^{ns}	.009
3. Thinking Dispositions Block: AOT; Super; CFC	.008	1.03 ^{ns}	.005
4. Cognitive Reflection Test	.064	27.68	.064
Overall Regression: $F = 10.44$			
Multiple $R = .468$			
Multiple $R^2 = .219$			
Criterion Variable = Rational-Thinking Composite			
1. Intelligence Block: WASI Vocab; WASI Matrix; WM	.227	33.52	.074
2. Thinking Dispositions Block: AOT; Super; CFC	.021	3.14	.016
3. Cognitive Reflection Test	.108	56.71	.108
Overall Regression: $F = 26.70$			
Multiple $R = .597$			
Multiple $R^2 = .356$			

WASI = Wechsler Abbreviated Scale of Intelligence; Set-S = set shifting; WM = working memory test; AOT = Actively Open-Minded Thinking scale; Super = superstitious thinking; CFC = Consideration of Future Consequences scale. ^{ns} Nonsignificant F value; all other F values reached significance

they are the last to be entered into the regression equation. This uniqueness value provides a comparative look at the potency of the four variable types as predictors, separate from the others. There we see that the CRT accounts for over twice as much unique variance (11.2% vs. 4.2%) as the next best predictor (the intelligence block).

Analyses of the individual components of the rational-thinking composite were largely parallel, with one or two notable exceptions. The next analysis is similar to the previous one, except that the criterion variable is the heuristics-and-biases task score. Each of the four blocks was statistically significant ($p < .001$) when entered

hierarchically. The CRT's ability to predict performance on this variable was again not due to variance shared with cognitive ability, executive functioning, or thinking dispositions. The CRT was once again the variable that predicted the most unique variance (8.0%), but in this analysis, the thinking dispositions block was the next most potent unique predictor (4.0%).

The next analysis is similar to the previous one, except that the criterion variable is the syllogistic reasoning score. Only Block 1 (intelligence) and Block 4 (the CRT) were significant ($p < .001$) when entered hierarchically, and only those two variable sets predicted unique variance (5.1% and

6.4%, respectively; $p < .001$). The CRT was once again the variable that predicted the most unique variance, but in this analysis the intelligence block predicted almost as much unique variance.

In the analyses completed so far, the CRT was a very potent predictor and intelligence a moderate predictor. The executive-functioning measures were not strong unique predictors in these analyses. However, Friedman et al. (2006) have shown that working memory tasks can be as strongly associated with cognitive ability as they are with other executive-functioning measures. Indeed, the zero-order correlations in Table 2 indicate that a cognitive ability measure (CA2) including working memory correlates more highly with rational-thinking performance than does the WASI alone. Thus, the final analysis in Table 3 groups the working memory task in the intelligence block for perhaps a fairer look at how strong a predictor intelligence is relative to the CRT.

The criterion variable in this final hierarchical regression analysis was the rational-thinking composite score. The first block of variables entered were the WASI Vocabulary, WASI Matrix, and working memory scores, and they accounted for 22.7% of the variance ($p < .001$). The second block of variables entered were the three thinking disposition measures, and they accounted for an additional 2.1% of the variance ($p < .05$). Finally, scores on the CRT were entered into the equation and accounted for a substantial amount of unique variance (10.8%, $p < .001$). The far right column indicates that the CRT was the more potent unique predictor of the three (10.8% unique variance vs. 7.4% and 1.6%).

As an additional way to reveal the overlap in the variables as predictors of rational thinking, we conducted a commonality analysis (Pedhazur, 1997) in which the variance explained by each variable was partitioned into a portion unique to that variable and portions shared with every possible combination of other variables. Table 4 presents a commonality analysis that displays the unique and over-

lapping variance of the CRT, the expanded cognitive ability block (WASI Vocabulary, WASI Matrix, and working memory scores), and the thinking disposition block in explaining performance on the rational-thinking composite. The first row indicates the unique variance in the rational-thinking composite explained by each of the predictors. The next row displays the explained variance in the rational-thinking composite that is common to the CRT and the cognitive ability block (10.2%). The third row displays the explained variance in the rational-thinking composite that is common to the CRT and the thinking dispositions block (0.5%). The fourth row displays the explained variance in the rational-thinking composite that is common to the cognitive ability block and the thinking dispositions block (2.8%). The fifth row indicates that the explained variance in the rational-thinking composite that is common to all three predictors is 2.3%. All of the variance components added together (.108 + .074 + .016 + .102 + .005 + .028 + .023) sum to the total variance explained in the rational-thinking composite score by the three groups of predictors: 35.6%.

Discussion

The CRT is moderately associated with both cognitive ability and rational-thinking skill. Its .49 correlation with the rational-thinking composite variable was the highest correlation of any predictor. Nonetheless, because the CRT also overlaps with cognitive ability, it is possible that it is through cognitive ability that it garners its predictive power. Several of the regression analyses reported indicated that this was not the case—that the CRT could predict rational-thinking performance independent not only of intelligence, but also of executive functioning and thinking dispositions. In fact, in all of the analyses in Table 3, the CRT accounted for more unique variance explained than did the block of intelligence measures.

Table 4 Results of a commonality analysis using the rational-thinking composite score as a criterion variable

	1. CRT	2. Cognitive Ability: Vocab; Matrix; WM	3. Thinking Dispositions: AOT; Super; CFC
Unique Variance	.108	.074	.016
Common 1 & 2	.102	.102	
Common 1 & 3	.005		.005
Common 2 & 3		.028	.028
Common 1 & 2 & 3	.023	.023	.023
Total Unique + Common Variance	.237	.227	.071

CRT = Cognitive Reflection Test; Vocab = Wechsler Abbreviated Scale of Intelligence (WASI) Vocabulary subscale; Matrix = WASI Matrix Reasoning subscale; WM = working memory test; AOT = Actively Open-Minded Thinking scale; Super = superstitious thinking; CFC = Consideration of Future Consequences scale; Unique Variance = explained variance in rational-thinking composite scores that is unique to that variable block; Common 1 & 2 = explained variance that is common to Variable Blocks 1 and 2; Common 1 & 3 = explained variance that is common to Variable Blocks 1 and 3; Common 2 & 3 = explained variance that is common to Variable Blocks 2 and 3; Common 1 & 2 & 3 = explained variance that is common to all three variable blocks.

The CRT also consistently predicted more variance in criterion variables than did the executive-functioning measures. Perhaps this is surprising, because doing well on the CRT would seem to stress the same set-shifting and inhibitory control features that have been emphasized in recent work on executive functioning (Aron, 2008; Best et al., 2009; Handley, Capon, Beveridge, Dennis, & Evans, 2004; Hasher et al., 2007; Miyake et al., 2000; Zelazo, 2004). It is possible that our executive-functioning measures were, as a group, too thin and heterogeneous. That is, we assessed working memory as well as set shifting and inhibition in the block of executive-functioning tasks, but we did so with only one task per construct (see Miyake et al., 2000, and Salthouse et al., 2003, for multiple-measures approaches). Perhaps if we had focused on inhibition and measured that construct with multiple tasks, we might have found more overlap between the executive-functioning construct and the CRT. Nonetheless, as operationalized in this study, we found that the CRT explains substantial variance in rational thinking that cannot be accounted for by our measures of cognitive ability, executive functioning, or thinking dispositions. What may be the reason for the surprisingly unique predictive power of the CRT?

It has only recently been fully recognized that intelligence and other cognitive ability tests leave out important domains of human cognition (Stanovich, 2009b). In psychology and among the lay public alike, assessments of intelligence and tests of cognitive ability are taken to be the sine qua non of good thinking. Critics of these instruments often point out that IQ tests fail to assess many domains of psychological functioning that are essential. For example, many largely noncognitive domains, such as socioemotional abilities, creativity, empathy, and interpersonal skills, are almost entirely unassessed by tests of cognitive ability. However, even these common critiques of intelligence tests often contain the unstated assumption that although intelligence tests miss certain key noncognitive areas, they encompass most of what is important cognitively. Recent work on individual differences in cognitive function has begun to challenge this assumption (Bruine de Bruin et al., 2007; Oechssler et al., 2009; Stanovich, 2009b, 2011; Stanovich & West, 2007, 2008a, 2008b).

That there is reliable variance in rational thinking independent of intelligence has been suggested before (Stanovich & West, 1998c, 2008b; West et al., 2008), but the properties of this intelligence-partialed variance are largely unexplored. The CRT appears to be a promising measure in this respect. Heuristics-and-biases tasks collectively measure a construct that we might term *rational thought*. Research has shown that there does appear to be reliable variance in rational thinking over and above what can be predicted by cognitive ability (Bruine de Bruin et al., 2007; Finucane & Gullion, 2010; Stanovich, 2011). The

CRT measures properties relevant to rational thinking that go beyond those measured on intelligence tests. That there is reliable variance in rational thinking independent of intelligence has been suggested before (Stanovich & West, 1998c, 2008b; West et al., 2008), but the properties of this intelligence-partialed variance are largely unexplored. The CRT appears to be a promising measure in this respect. We have shown here that the CRT can explain a substantial amount of this reliable variance. In order to determine why this is the case, it might be useful to think in terms of a classification scheme for rational-thinking errors discussed by Stanovich, Toplak, and West (2008; see Stanovich, 2009b, 2011). Their taxonomy is based around the finding that the human brain has two broad characteristics that make it less than rational. One is a processing problem and one a content problem, and intelligence provides insufficient inoculation against both.

The processing problem is the one mentioned in our introductory discussion: that humans tend to be cognitive misers. This has been a major theme throughout the past 40 years of research in the cognitive science of human judgment and decision making (Dawes, 1976; Simon, 1955, 1956; Taylor, 1981; Tversky & Kahneman, 1974). For example, Kahneman and Frederick (2002) discuss attribute substitution as a common mechanism used to lighten cognitive load. Attribute substitution occurs when a person needs to assess attribute A but finds that assessing attribute B (which is correlated with A) is easier cognitively, and so uses B instead. In simpler terms, attribute substitution amounts to substituting an easier question for a harder one.

Humans are cognitive misers because their basic tendency is to default to heuristic processing mechanisms of low computational expense. This bias to default to the simplest cognitive mechanism, however, means that humans are sometimes less than rational. Heuristic processes often provide a quick solution that is a first approximation to an optimal response. But modern life often requires more precise thought than this. Modern technological societies are in fact hostile environments for people reliant on only the most easily computed automatic response (Stanovich, 2009b, 2011). Thus, being cognitive misers will sometimes impede people from achieving their goals. Many effects in the heuristics-and-biases literature are the results of the human tendency to default to miserly processing: anchoring biases, framing effects, preference reversals, nondisjunctive reasoning, myside biases, and status quo biases, to name just a few.

The second broad reason that humans are less than rational represents a content problem. Normative responding on a cognitive task often requires that responses based on heuristic processing be overridden and replaced by responses that are more accurately computed (Evans, 2003, 2008, 2010; Evans & Frankish, 2009). However, the override process is not simply procedural but instead utilizes content—that is, it

uses declarative knowledge and strategic rules (linguistically coded strategies). Gaps in these knowledge structures represent a second major class of reasoning error. If one is going to trump a heuristic response with conflicting information or a learned rule, one must have previously learned the information or the rule. Rational-thinking errors due to such knowledge gaps can occur in a potentially large set of coherent knowledge bases in the domains of probabilistic reasoning, causal reasoning, logic, and scientific thinking (the importance of alternative hypotheses, etc.).

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. In contrast, the CRT does seem highly relevant to the idea of humans as cognitive misers. As mentioned in the introduction, the CRT is unlike traditional insight tasks in the reasoning literature (Gilhooly & Fioratou, 2009). Insight problems are not failed because the participant fails to think enough; often, they spend minutes immersed in intense thought, but nonetheless fail to derive the correct solution. In traditional insight problems (e.g., the nine-dot problem), participants spend a long time thinking because no viable solution at all occurs to them. The type of error made on the CRT is different. On this test, an incorrect answer is initially primed. However, miserly processing ensures that it is not overridden and replaced by a superior response.

Interpreted in this way, the CRT becomes in part a measure of rational thought, rather than a distal predictor or an underlying ability supporting rational thought. This type of interpretation is consistent with its high correlation with the rational-thinking composite score. In short, the CRT is a measure of the tendency toward the class of reasoning error that derives from miserly processing. This may be why the predictive power of the CRT is in part separable from cognitive ability. Intelligence tests do not assess the tendency toward miserly processing in the way that the CRT does. Instead, the former measures computational power that is available to the participant, but not necessarily the depth of processing that is typically used in most situations. In fact, the CRT might be a particularly potent measure of miserly tendencies because of its logic of construction: It is a performance measure¹ rather than a self-report measure. That is, it is not a questionnaire measure on which people indicate their preferences for engagement—for example, as

the need-for-cognition scale does (Cacioppo, Petty, Feinstein, & Jarvis, 1996). Instead, the tendency to accept heuristically triggered responses is measured in a real performance context where participants are searching for an accurate solution. The CRT measures miserliness in action, so to speak. It is a direct measure of miserly processing rather than an indirect self-report indicator.

Appendix: Descriptions of individual heuristics-and-biases tasks

Causal base rate

In this problem, adapted from Fong, Krantz, and Nisbett (1986), a couple are deciding to buy one of two otherwise equal cars. Preference for the opinion of experts and the large-sample information over salient personal testimony was scored as 1 (the reverse was scored as 0).

Sample size: Hospital problem

This problem was the classic sample-size problem studied by Tversky and Kahneman (1974).

Sample size: Squash problem

This problem was taken from Kahneman and Tversky (1982). Participants were told that a game of squash can be played to either 9 or 15 points. Holding all other rules of the game constant, if A is a better player than B, the participants are asked which scoring scheme would give player A a better chance of winning. Like the hospital problem, this item is used to explore participants' understanding that, other things being equal, a larger sample size is more likely to approximate a population value. In this case, the better player's chances of winning would increase when there are more scoring opportunities, and the 15-point scoring system is the correct choice.

Regression to the mean

Drawn from Lehman, Lempert, and Nisbett (1988), this problem was worded as follows:

After the first 2 weeks of the major league baseball season, newspapers begin to print the top 10 batting averages. Typically, after 2 weeks, the leading batter often has an average of about .450. However, no batter in major league history has ever averaged .450 at the end of the season. Why do you think this is? Circle one:

- a. When a batter is known to be hitting for a high average, pitchers bear down more when they pitch to him.

¹ One astute reviewer mentioned the caution that in order to remain a performance measure rather than a self-report measure, the familiarity of the CRT items will need to be assessed. That is, we must be cautious about the growing publicity that the CRT is receiving. Clearly, if individuals become familiar with the items, the CRT can no longer be considered a performance measure. Most studies try to assess whether participants have seen the problems before. The ultimate answer will be in the generation of more CRT items that vary in their surface characteristics.

- b. Pitchers tend to get better over the course of a season, as they get more in shape. As pitchers improve, they are more likely to strike out batters, so batters’ averages go down.
- c. A player’s high average at the beginning of the season may be just luck. The longer season provides a more realistic test of a batter’s skill.
- d. A batter who has such a hot streak at the beginning of the season is under a lot of stress to maintain his performance record. Such stress adversely affects his playing.
- e. When a batter is known to be hitting for a high average, he stops getting good pitches to hit. Instead, pitchers “play the corners” of the plate because they don’t mind walking him.

Response c is the only response that shows some recognition of the possibility of regression effects, and was scored as 1, while the other options were scored as 0.

Gambler’s fallacy 1

In the first gambler’s fallacy problem, the slot machine problem, the participant read the following: “When playing slot machines, people win something about 1 in every 10 times. Julie, however, has just won on her first three plays. What are her chances of winning the next time she plays? ____ out of ____.” The correct response, 1 out of 10, was scored as correct, and all other responses were scored as incorrect.

Gambler’s fallacy 2

In the second gambler’s fallacy problem, the coin problem, the participant read the following:

Imagine that we are tossing a fair coin (a coin that has a 50/50 chance of coming up heads or tails) and it has just come up heads 5 times in a row. For the 6th toss do you think that:

- a. It is more likely that tails will come up than heads.
- b. It is more likely that heads will come up than tails.
- c. Heads and tails are equally probable on the sixth toss.

Answer c is the correct response and was scored as 1, while the other two alternatives were scored as 0.

Conjunction problem

This problem was based on Tversky and Kahneman’s (1983) much-studied Linda problem. Responses indicating that the conjunction was more likely than one of its components were incorrect and scored as 0, and all other responses were scored as 1.

Covariation detection

This problem appeared as follows:

A doctor had been working on a cure for a mysterious disease. Finally, he created a drug that he thinks will cure people of the disease. Before he can begin to use it regularly, he has to test the drug. He selected 300 people who had the disease and gave them the drug to see what happened. He selected 100 people who had the disease and did not give them the drug in order to see what happened. The table below indicates what the outcome of the experiment was:

	Cure	
	Yes	No
Treatment present	200	100
Treatment absent	75	25

Participants were asked to judge whether this treatment was positively or negatively associated with the cure for this disease by circling a number from a scale ranging from -10 (strong negative association) to +10 (strong positive association). Negative judgments, which indicated the inefficacy of the treatment, were scored as correct.

Methodological reasoning

Adapted from the Middleton problem of Lehman et al. (1988), this multiple-choice problem has only one alternative that indicates the ability to reason methodologically about confounded variables in everyday life. This alternative was scored as 1, and the other responses as 0.

Bayesian reasoning

This problem was the David Maxwell problem, adapted from Beyth-Marom and Fischhoff (1983) and studied by Stanovich and West (1998d). It is used to assess Bayesian belief updating.

Framing problem

A fundamental assumption of decision theory is that of descriptive invariance: “that the preference order between prospects should not depend on the manner in which they are described” (Kahneman & Tversky, 1984, p. 343). The disease problem of Tversky and Kahneman (1981) is a classic problem in which participants sometimes do not display descriptive invariance. Instead, they display a framing effect. This problem is presented in two parts (within subjects), positive and negative framing. Descrip-

tive invariance was correct, and scored as 1. Violation of description invariance was scored as 0.

Probabilistic reasoning: Denominator neglect

This probabilistic reasoning task was a marble game that was modeled on a task introduced by Kirkpatrick and Epstein (1992; see also Denes-Raj & Epstein, 1994; Reyna, 1991; Reyna & Brainerd, 1994, 2008). The problem read as follows:

Assume that you are presented with two trays of black and white marbles: a large tray that contains 100 marbles and a small tray that contains 10 marbles. The marbles are spread in a single layer on each tray. You must draw out one marble (without peeking, of course) from either tray. If you draw a black marble, you win \$2. Consider a condition in which the small tray contains 1 black marble and 9 white marbles, and the large tray contains 8 black marbles and 92 white marbles. [A drawing of two trays with their corresponding numbers of marbles arranged neatly in 10-marble rows appeared above the previous sentence.] From which tray would you prefer to select a marble in a real situation?

The correct response was the small tray, because the chance of pulling a black marble was 10% from the small tray, whereas the chance of pulling a winning marble was 8% from the large tray.

Probability matching

This problem was the dice problem adapted from West and Stanovich (2003; see also Gal & Baron, 1996). Students who preferred Strategy D of predicting “red” for each of the 60 rolls were classified as using the maximizing strategy, which was scored as correct. All other strategies were scored as incorrect.

Sunk cost

This problem from Frisch (1993; see Stanovich & West, 1998b) was the movie problem, which has two parts. In the first part, participants are told to imagine that they are staying in a hotel room, and they have just paid \$6.95 to see a movie on pay TV. Then they are told that they are bored 5 min into the movie and that the movie seems pretty bad. They are then asked whether they would continue to watch the movie or switch to another channel. In the second part, the scenario is analogous, except that they have not had to pay for the movie. They are asked again whether they would continue to watch the movie or switch to another channel. Responses were scored as correct if the participant

consistently chose across the two situations (either continuing to watch the movie in both cases, or switching to another channel in both cases), and as incorrect if the participant displayed a sunk cost (that is, continuing to watch the movie if it had been paid for but not if it was free).

Outcome bias

Our measure of outcome bias derived from a problem investigated by Baron and Hershey (1988), and was composed of two parts presented separately in the battery of measures. In Part 1, participants were told about a 55-year-old man who had a heart condition and whose operation succeeded. The probability of mortality from surgery was 8%. Participants responded on a 7-point scale ranging from 1 (*incorrect, a very bad decision*) to 7 (*clearly correct, an excellent decision*). Later, in the battery for Part 2 of this problem, participants evaluated a different decision to perform surgery on a patient with a hip condition that was designed to be objectively better than the first (2% chance of death rather than 8%), even though it had an unfortunate negative outcome (death of the patient). If participants rate the decision on the positive outcome case as better than the negative outcome decision, then they have displayed outcome bias. The absence of outcome bias was scored as the correct response for this problem.

References

- Ackerman, P. L. (1994). Intelligence, attention, and learning: Maximal and typical performance. In D. K. Detterman (Ed.), *Current topics in human intelligence* (Vol. 4, pp. 1–27). Norwood: Ablex.
- Ackerman, P. L. (1996). A theory of adult development: Process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257.
- Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.
- Ackerman, P. L., & Kanfer, R. (2004). Cognitive, affective, and conative aspects of adult intellect within a typical and maximal performance framework. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 119–141). Mahwah: Erlbaum.
- Aron, A. R. (2008). Progress in executive-function research: From tasks to functions to regions to networks. *Current Directions in Psychological Science*, *17*, 124–129.
- Baron, J. (2008). *Thinking and deciding* (4th ed.). New York: Cambridge University Press.
- Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569–579.
- Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*, 180–200.
- Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, *45*, 1185–1195.

- Bishop, M. A., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment*. Oxford: Oxford University Press.
- Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 92*, 938–956.
- Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197–253.
- Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making, 5*, 182–191.
- 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*, 20–33.
- Cronbach, L. J. (1949). *Essentials of psychological testing*. New York: Harper.
- Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 3–11). Hillsdale: Erlbaum.
- Del Missier, F., Mantyla, T., & Bruine de Bruin, W. (2010). Executive functions in decision making: An individual differences approach. *Thinking & Reasoning, 16*, 69–97.
- Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology, 66*, 819–829.
- De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition, 106*, 1248–1299.
- Dohmen, T., Falk, A., Huffman, D., Marklein, F., & Sunde, U. (2009). Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample. *Journal of Economic Behavior and Organization, 72*, 903–915.
- Duckworth, A. L. (2009). Over and beyond high-stakes testing. *American Psychologist, 64*, 279–280.
- Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., et al. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General, 137*, 131–148.
- Epstein, S., & Meier, P. (1989). Constructive thinking: A broad coping variable with specific components. *Journal of Personality and Social Psychology, 57*, 332–350.
- Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 451–468.
- Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454–459.
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology, 59*, 255–278.
- Evans, J. St. B. T. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.
- Evans, J. St. B. T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*, 295–306.
- Evans, J. St. B. T., & Frankish, K. (Eds.). (2009). *In two minds: Dual processes and beyond*. Oxford: Oxford University Press.
- Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology Press.
- Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging, 25*, 271–288.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253–292.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42.
- Friedman, N. P., Haberstic, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., et al. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science, 18*, 893–900.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*, 172–179.
- Frisch, D. (1993). Reasons for framing effects. *Organizational Behavior and Human Decision Processes, 54*, 399–429.
- Gal, I., & Baron, J. (1996). Understanding repeated simple choices. *Thinking & Reasoning, 2*, 81–98.
- George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86*, 93–111.
- Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking and Reasoning, 15*, 355–376.
- Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning, 11*, 279–302.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Gonzalez, R., Grant, I., Miller, W., Taylor, M. J., Schweinsburg, B. C., Carey, C. L., et al. (2006). Demographically adjusted normative standards for new indices of performance on the Paced Auditory Serial Addition Task (PASAT). *Clinical Neuropsychologist, 20*, 396–413.
- Gronwall, D. M. A. (1977). Paced auditory serial-addition task: A measure of recovery from concussion. *Perceptual and Motor Skills, 44*, 367–373.
- Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J St B T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning, 10*, 175–195.
- Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York: Oxford University Press.
- Jones, W., Russell, D., & Nickel, T. (1977). Belief in the paranormal scale: An objective instrument to measure belief in magical phenomena and causes. *JSAS Catalog of Selected Documents in Psychology, 7*(100), Ms. No. 1577.
- Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences, 23*, 681–683.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11*, 123–141.
- Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist, 39*, 341–350. doi:10.1037/0003-066X.39.4.341.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582–591.
- Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.
- Kirkpatrick, L., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Evidence for two conceptual systems. *Journal of Personality and Social Psychology, 63*, 534–544.
- Klaczynski, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision making. *Child Development, 72*, 844–861.

- Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition*, *38*, 667–676.
- Kokis, J., Macpherson, R., Toplak, M., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.
- Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effect of graduate training on reasoning. *American Psychologist*, *43*, 431–442.
- Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259–289.
- Lieberman, M. D. (2009). What zombies can't do: A social cognitive neuroscience approach to the irreducibility of reflective consciousness. In J. St B T Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 293–316). Oxford: Oxford University Press.
- Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science & myth*. Cambridge: MIT Press.
- Miyake, A., Friedman, N., Emerson, M. J., & Witzki, A. H. (2000). The utility and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove: Midwest Publications.
- Obrecht, N. A., Chapman, G. B., & Gelman, R. (2009). An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition*, *37*, 632–643.
- Oechssler, J., Roeder, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*, 147–152.
- Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual differences approach. *Journal of Behavioral Decision Making*, *18*, 1–27.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth: Harcourt Brace.
- Perkins, D. N. (1995). *Outsmarting IQ: The emerging science of learnable intelligence*. New York: Free Press.
- Reitan, R. M. (1955). The relation of the Trail Making Test to organic brain damage. *Journal of Consulting Psychology*, *19*, 393–394. doi:10.1037/h0044509.
- Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.
- Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review*, *11*, 317–336.
- Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239–272). New York: Wiley.
- Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*, 89–107.
- Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative theory of judgment and decision making. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 201–245). New York: Cambridge University Press.
- Sá, W., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *91*, 497–510.
- Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*, 482–486.
- Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*, 566–594.
- Samuels, R., & Stich, S. P. (2004). Rationality and psychology. In A. R. Mele & P. Rawling (Eds.), *The Oxford handbook of rationality* (pp. 279–300). Oxford: Oxford University Press.
- Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, *53*, 491–517.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
- Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). New York: Cambridge University Press.
- Slugoski, B. R., Shields, H. A., & Dawson, K. A. (1993). Relation of conditional reasoning to heuristic processing. *Personality and Social Psychology Bulletin*, *19*, 158–166.
- Stanovich, K. E. (1989). Implicit philosophies of mind—The dualism scale and its relation to religiosity and belief in extrasensory perception. *Journal of Psychology*, *123*, 5–23.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah: Erlbaum.
- Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
- Stanovich, K. E. (2008). Higher-order preferences and the master rationality motive. *Thinking & Reasoning*, *14*, 111–127.
- Stanovich, K. E. (2009a). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford: Oxford University Press.
- Stanovich, K. E. (2009b). *What intelligence tests miss: The psychology of rational thought*. New Haven: Yale University Press.
- Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.
- Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristics and biases. *Advances in child development and behavior*, *36*, 251–285.
- Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*, 342–357.
- Stanovich, K. E., & West, R. F. (1998a). Cognitive ability and variation in selection task performance. *Thinking and Reasoning*, *4*, 193–230.
- Stanovich, K. E., & West, R. F. (1998b). Individual differences in framing and conjunction effects. *Thinking and Reasoning*, *4*, 289–317.
- Stanovich, K. E., & West, R. F. (1998c). Individual differences in rational thought. *Journal of Experimental Psychology: General*, *127*, 161–188.
- Stanovich, K. E., & West, R. F. (1998d). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition*, *26*, 161–179.
- Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology*, *38*, 349–385.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–726.
- Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, *13*, 225–247. doi:10.1080/13546780600780796.
- Stanovich, K. E., & West, R. F. (2008a). On the failure of intelligence to predict myside bias and one-sided bias. *Thinking & Reasoning*, *14*, 129–167.

- Stanovich, K. E., & West, R. F. (2008b). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*, 672–695.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Intelligence and rationality. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence*. New York: Cambridge University Press.
- Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. Cambridge: Cambridge University Press.
- Strathman, A., Gleicher, F., Boninger, D. S., & Scott Edwards, C. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*, 742–752.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests* (3rd ed.). New York: Oxford University Press.
- Taylor, S. E. (1981). The interface of cognitive and social psychology. In J. H. Harvey (Ed.), *Cognition, social behavior, and the environment* (pp. 189–211). Hillsdale: Erlbaum.
- Toplak, M., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (2007). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making, 20*, 103–124.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI)*. San Antonio: Harcourt Brace, Psychological Corp.
- West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition, 31*, 243–251.
- West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*, 930–941.
- Zeidner, M., & Matthews, G. (2000). Intelligence and personality. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 581–610). New York: Cambridge University Press.
- Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences, 8*, 12–17.

Author Note

This research was supported by grants from the Social Sciences and Humanities Research Council of Canada to M.E.T. and the Canada Research Chairs program to K.E.S.