Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ptar20

Assessing miserly information processing: An expansion of the Cognitive Reflection Test

Maggie E. Toplak[a], Richard F. West[b] & Keith E. Stanovich[c]

[a] Department of Psychology, York University, Toronto, Canada

[b] Department of Graduate Psychology, James Madison University, Harrisonburg, VA, USA

[c] Department of Applied Psychology and Human Development, University of Toronto, Toronto, Canada

Published online: 28 Oct 2013.

To cite this article: Maggie E. Toplak, Richard F. West & Keith E. Stanovich , Thinking & Reasoning (2013): Assessing miserly information processing: An expansion of the Cognitive Reflection Test, Thinking & Reasoning, DOI: 10.1080/13546783.2013.844729

To link to this article: http://dx.doi.org/10.1080/13546783.2013.844729

Thinking & Reasoning, 2013
http://dx.doi.org/10.1080/13546783.2013.844729

Routledge
Taylor & Francis Group

Assessing miserly information processing: An expansion of the Cognitive Reflection Test

Maggie E. Toplak[1], Richard F. West[2], and Keith E. Stanovich[3]

[1]Department of Psychology, York University, Toronto, Canada
[2]Department of Graduate Psychology, James Madison University, Harrisonburg, VA, USA
[3]Department of Applied Psychology and Human Development, University of Toronto, Toronto, Canada

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. It is a prime measure of the miserly information processing posited by most dual process theories. The original three-item test may be becoming known to potential participants, however. We examined a four-item version that could serve as a substitute for the original. Our data show that it displays a .58 correlation with the original version and that it has very similar relationships with cognitive ability, various thinking dispositions, and with several other rational thinking tasks. Combining the two versions into a seven-item test resulted in a measure of miserly processing with substantial reliability (.72). The seven-item version was a strong independent predictor of performance on rational thinking tasks after the variance accounted for by cognitive ability and thinking dispositions had been partialled out.

Keywords: Cognitive Reflection Test; Rational thinking; Cognitive ability; Thinking dispositions; Dual process theory.

One background assumption of most dual process theories is that people tend to be cognitive misers in their thinking. This is what makes the override function

Correspondence should be addressed to Maggie E. Toplak, Department of Psychology, York University, 126 BSB, 4700 Keele St. Toronto, M3J 1P3, Ontario Canada. E-mail: mtoplak@yorku.ca

Preparation of this manuscript was supported by grants from the Social Sciences and Humanities Research Council of Canada to Maggie E. Toplak and from the John Templeton Foundation to Keith E. Stanovich and Richard F. West. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. The authors wish to thank Geoff Sorge, Mohamed Al-Haj, and Amanda Edwards for assistance with data collection and scoring the WASI.

in most dual process theories so important. The cognitive miser assumption that is retained in most modern dual process theories (Evans, 2008; Evans & Stanovich, 2013; Stanovich, 2004) has been a major theme throughout the past 50 years of research in psychology and cognitive science (Dawes, 1976; Evans, 2010; Johnson-Laird, 1983, 1999; Kahneman, 2011; Simon, 1955, 1956; Stanovich, 2009; Taylor, 1981; Tversky & Kahneman, 1974).

When approaching any problem, our brains have available various computational mechanisms for dealing with the situation. These mechanisms embody a trade-off, however, well described in contemporary dual process theory (Evans & Stanovich, 2013). The defining feature of Type 1 processing is its autonomy—their execution is mandatory when the triggering stimuli are encountered and they can operate in parallel without interfering with themselves or with Type 2 processing. Type 2 processing is relatively slow and computationally expensive, and one of its most critical functions is to override Type 1 processing.

The trade-off between Type 1 and Type 2 processing is one between power and expense. Type 2 processing enables us to solve a wide range of novel problems, and solve them with great accuracy. However, this power comes with a cost. Type 2 processing takes up a great deal of attention, tends to be slow, tends to interfere with other thoughts and actions that we are carrying out, and requires great concentration that is often experienced as aversive. In contrast, Type 1 processes are low in computational power but have the advantage that they are low in cost. These mechanisms cannot solve a wide range of problems and do not permit fine-grained accuracy, but they are fast acting, do not interfere with other ongoing cognition, require little concentration, and are not experienced as aversive.

Humans are cognitive misers because their basic tendency is to default to Type 1 processing mechanisms of low computational expense. Using less computational capacity for one task means that there is more left over for another task if they both must be completed simultaneously. This would seem to be adaptive. Nevertheless, this strong bias to default to the simplest cognitive mechanism—to be a cognitive miser—means that humans are often less than rational. Type 1 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 (see Kahneman, 2011; Stanovich, 2004; Sunstein, 2013; Thaler & Sunstein, 2008).

Because being a cognitive miser will seriously impede people from achieving their goals, psychologists have been interested in studying individual differences in the miserly tendency (Stanovich, 1999; Stanovich & West, 2000). People vary in how likely they are to override a prepotent response

alternative that is incorrect and to engage in further reflection that leads to the correct response. By far the most popular measure of miserly processing has been Frederick's (2005) Cognitive Reflection Test (CRT). The problems on the CRT seem at first glance to be similar to the well-known insight problems in the problem-solving literature, but in fact they display a critical difference. Classic insight problems (see Gilhooly & Fioratou, 2009) do not usually trigger an attractive alternative response. Instead the participant sits lost in thought trying to reframe the problem correctly as in, for example, the classic nine dot problem. The three problems on the CRT are of interest to researchers working in the dual-process tradition because a strong alternative response is initially primed and then must be overridden.

Shockingly, since it is based on just three items, the CRT has proven to be a potent predictor of performance on rational thinking tasks. Frederick (2005) observed that his CRT could predict the tendency to choose high expected-value gambles and that CRT scores were associated with temporal discounting and framing. Likewise, Cokely and Kelley (2009) found a correlation between performance on the CRT and the proportion of choices consistent with expected value. Others have found the CRT to be significantly associated with avoiding the conjunction fallacy; expected value choices; maximising strategies on probabilistic prediction tasks; the endorsement of profit maximising strategies; the avoidance of the illusion of explanatory depth; non-superstitious thinking; performance calibration, and general numeracy (Fernbach et al., 2013; Koehler & James, 2010; Liberali et al., 2012; Mata et al., 2013; Moritz et al., 2013; Oechssler et al., 2009; Pennycook et al., 2012; Shenhav et al., 2012).

In the most comprehensive study yet, Toplak, West, and Stanovich (2011) formed a composite variable of 15 separate rational thinking tasks from many different domains in the heuristics and biases literature. They found that the CRT was a better predictor of rational thinking than either measures of intelligence or measures of executive functioning. Several of the regression analyses conducted indicated that the CRT could predict rational thinking performance independent of not only intelligence but also executive functioning and thinking dispositions. In fact, in all of the analyses, the CRT by itself accounted for more unique variance explained than the block of cognitive ability measures (intelligence). This is astounding predictive performance for a three-item measure!

Nevertheless, there are problems on the horizon for the CRT going into the future. The items are becoming extremely well known—especially the famous bat-and-ball item. The latter is used in countless classroom demonstrations now, and it has appeared in many magazines and famous books—most notably Daniel Kahneman's rightly lauded and extensively reviewed *Thinking, Fast and Slow* (2011). From the standpoint of reliability, three

items is obviously too few. Finally, in some populations, the overall score on the three-item version might be floored. Frederick (2005) reported the mean performance on the three items across a variety of academic institutions and found that, for example, students at Michigan State University and Bowling Green State University got less than one item out of three correct. The mean for the University of Toledo was just 0.57. Clearly, using the three-item version in high schools and community colleges will be problematic in terms of floor effects.

Thus the CRT is badly in need of supplement and extension. Here we report the results of using a seven-item CRT, one that includes the original three items reported by Frederick (2005) and four others without the extensive research track-record of the original problems. We examined its ability to predict performance on seven rational thinking tasks from the heuristics and biases literature and whether the four new items add to the variance explained. In order to situate the seven-item version within the overall space of individual differences, we also assessed cognitive ability (intelligence and executive functioning) and four different thinking dispositions (Need for Cognition, Actively Openminded Thinking, Superstitious Thinking, and Consideration of Future Consequences).

METHOD

Participants and procedure

A total of 160 participants (M age = 20.7 years, SD = 3.7; 63 males and 97 females) took part in the study. The participants were recruited at a large university and were either part of a participant pool who received course credit (n = 123) or paid (n = 37) for their participation. The paid participants were older than the unpaid participants (M difference = 3.4 years); $t(158) = 5.15$, $p < .001$, but did not differ significantly from the unpaid participants in sex, cognitive ability (WASI) scores, high school GPA, or college GPA. Participants provided estimations of their current university grade-point averages using the university's percentage scale (M = 73.3%, SD = 7.8; On this university's grading scale, 70–74% corresponds to a B letter grade).

Participants completed the battery of tasks described below plus some other measures during a single, 2-hour session. The tasks were presented in the following order: WASI, demographics part 1, otherside thinking tasks, framing problems, denominator neglect, belief bias syllogistic reasoning, selection tasks, cognitive reflection test, bias blind spot, temporal discounting, thinking disposition measures, demographics part 2. All of the tasks were presented on a computer using MediaLab v2008 software, with the exception of the cognitive ability testing (WASI), which was administered individually by an examiner.

Tasks and variables

Cognitive Reflection Test (CRT)

Taken from Frederick (2005), the original test was composed of three questions, as follows:

(1) 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 [Correct answer = 5 cents; intuitive answer = 10 cents]

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes [Correct answer = 5 minutes; intuitive answer = 100 minutes]

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days [Correct answer = 47 days; intuitive answer = 24 days]

The score on these original three items will be designated CRT3 in our study.

We added the follow four-items to the CRT:

(4) If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? _____ days [correct answer = 4 days; intuitive answer = 9]

(5) Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? _____ students [correct answer = 29 students; intuitive answer = 30]

(6) A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? _____ dollars [correct answer = $20; intuitive answer = $10]

(7) Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money [correct answer = c, because the value at this point is $7,000; intuitive response = b].

Items #4 and #5 were kindly supplied to us by Shane Frederick in personal correspondence (13 June 2011); item #6 was adapted from Dominowski (1994; see Gilhooly & Murphy, 2005); and the seventh item was created by the authors. The score on these four new items will be designated CRT4 in our study. The score on all seven items will be labelled CRT7.

Cognitive ability: Wechsler abbreviated scales of intelligence

The Vocabulary and Matrix Reasoning subtests from the Wechsler Abbreviated Scales of Intelligence (WASI; Wechsler, 1999) were used as indices of verbal and nonverbal ability. These subtests were administered individually by an examiner. The mean raw score on the Vocabulary subtest was 56.3 ($SD = 6.9$), and the mean raw score of the Matrix Reasoning subtest was 26.0 ($SD = 4.2$). 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.

Thinking disposition measures

Participants completed a self-report questionnaire in which they were asked to rate their agreement with each question using the following six-point scale: Strongly Disagree (1), Disagree Moderately (2), Disagree Slightly (3), Agree Slightly (4), Agree Moderately (5), Strongly Agree (6). Questions were presented in mixed order so that the target scales of interest would be less transparent to participants. Several scales were intermixed in the questionnaire.

Need for cognition (NFC). This 18-item scale assesses the motive to engage in effortful cognitive activities (Cacioppo et al., 1996). Sample items include: "The notion of thinking abstractly is appealing to me", and "I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought". The mean score was $M = 68.6$ ($SD = 10.4$). The split-half reliability (Spearman-Brown corrected) of the need for cognition scale was .78 and Cronbach's alpha was .79.

Actively openminded thinking (AOT). This 41-item measure is scored in the direction that higher scores represented a greater tendency towards openminded thinking (Stanovich & West, 1997, 2007). 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 ($M = 164.0$, $SD = 21.3$). The split-half reliability (Spearman-Brown corrected) of the scale was .85 and Cronbach's alpha was .86.

Superstitious thinking (ST). This 13-item 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 include: "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 ($M = 34.2$, $SD = 10.9$). That score was turned into a z-score and the z-score reflected (multiplied by –1) so that higher scores indicate great resistance to superstitious thinking. Thus the superstitious thinking scale (reflected) would be expected, based on previous research, to correlate positively with the other thinking dispositions and with cognitive ability. The split-half reliability (Spearman-Brown corrected) of the scale was .86 and Cronbach's alpha was .84.

Consideration of future consequences (CFC). This 12-item scale assesses the extent to which individuals consider distant outcomes when choosing their present behaviour (Strathman, Gleicher, Boninger, & Edwards, 1994). 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 ($M = 49.3$, $SD = 7.4$). The split-half reliability (Spearman-Brown corrected) of the scale was .76 and Cronbach's alpha was .74.

Rational thinking tasks

Belief bias in syllogistic reasoning. Eight syllogistic reasoning problems, largely drawn from Markovits and Nantel (1989), were completed by the participants. Each problem was worded such that the validity judgement was in conflict with the believability of the conclusion. There were two types of these so-called inconsistent syllogisms. One type of inconsistent syllogism had a believable conclusion but an invalid format (e.g., "Premises: All living things need water; Roses need water; Conclusion: Roses are living things"—which is invalid). The other type had an unbelievable conclusions in a logically valid format (e.g., "Premises: All things that are smoked are good for the health; Cigarettes are smoked; Conclusion: Cigarettes are good for the health"—which is valid). Therefore the believability of the content was inconsistent with the logical format of the syllogism in both types. Problems of this type have typically been thought to mirror the critical thinking skill of being able to put aside one's prior knowledge and reason from new premises. After each item, the participants indicated their responses by selecting one of the two alternatives: (1) Conclusion follows logically from premises, or (2) Conclusion does not follow logically from premises. The eight syllogisms were presented together in the battery. A composite score of performance on the eight items was formed by summing the number of correct responses ($M = 3.86$, $SD = 1.5$); the split-half reliability (Spearman-Brown corrected) of the scale was .51 and Cronbach's alpha = .64.

Selection task. Three versions of the selection task were utilised, two with non-deontic content (Abstract and Destination Problem), and one with deontic content (Sears Problem). One non-deontic, abstract version was originally used by Wason (1966), and has been studied extensively in the reasoning literature. The second nondeontic problem was the Destination

Problem studied by Stanovich and West (1998a). The deontic version was the Sears Problem (Dominowski, 1995; Stanovich & West, 1998a). Each version of the selection task was separated in the battery by other rational thinking tasks. Because the rule is in the form of an *if P, then Q* rule, the participant must turn over the cards that could potentially falsify the rule—the P and not-Q cards, which was scored as the correct responses (and scored as 1). Because correct responding on nondeontic versions is typically so low, we also scored P-only choosers as correct (see Toplak & Stanovich, 2002), an alternative task construal championed by Margolis (1987). He has argued that turning the P card only is an appropriate response if the participant has adopted a so-called "open" reading of the rule—one where the cards represent classes rather than individual exemplars. All other selections were scored as 0. The scores on the three selection tasks were summed to form a selection task composite score ($M = 0.97$, $SD = 1.0$).

Denominator neglect. The five-problems on this task were modelled on Kirkpatrick and Epstein (1992; see also Denes-Raj & Epstein, 1994). An example of a trial read as follows:

> Assume that you are presented with two trays of black and white marbles (pictured below and right): The large tray contains 100 marbles. The small tray contains 10 marbles. The marbles are spread in a single layer in each tray. You must draw out one marble (without peeking, of course) from either tray. If you draw a black marble you win $5. Consider a condition in which: The small tray contains 1 black and 9 white marbles. The large tray contains 8 black and 92 white marbles. [A drawing of two trays with their corresponding numbers of marbles arranged neatly in 10-marbles-rows was pictured. The corresponding number of black and white marbles was printed in parentheses directly underneath each tray.] From which tray would you prefer to select a marble in a real situation?

The following scale was used to indicate preferences: (1) I would definitely pick from the small tray; (2) I would pick from the small tray; (3) I would probably pick from the small tray; (4) I would probably pick from the large tray; (5) I would pick from the large tray; (6) I would definitely pick from the large tray.

In the remaining four trials, the ratio of black:white numbers were as follows: 1:4 versus 19:81, 1:19 versus 4:96, 2:3 versus 19:31, and 3:12 versus 18:82. Each problem was separated in the battery by other rational thinking tasks. In all cases the correct response was to select the small tray, as the chances of pulling a black marble was higher (10% in the current example) than in the large tray (8%). The sum of the ratings of the five problems was reflected to form a composite score where higher values indicated more resistance to denominator neglect. The mean composite score across the five problems was 20.91 ($SD = 5.7$).

Temporal discounting. This five-item measure was adapted from Frederick (2005). For each item, participants indicated the strength of their preference for either a smaller amount of money now or a larger amount of money later. In each case the delayed larger amount corresponded to a substantial percentage increase in value, which on a simple interest basis would have resulted in value increases of between 40% to 240% if earned annually. The first item of this measure, for example, asked participants to indicate whether they would "prefer $3400 this month or $3800 next month". In this example, a willingness to wait was worth an extra $400—the equivalent of about an 11.8% gain in value in one month, which on a simple interest basis would have resulted in a value increase of about 141% if earned annually. Participants indicated their preferences using the following response scale: (1) I strongly prefer $3400 this month; (2) I slightly prefer $3400 this month; (3) I prefer $3400 this month; (4) I prefer $3800 next month; (5) I slightly prefer $3800 next month; (6) I strongly prefer $3800 next month. The remaining four items were the following: "$100 now or $140 next year"; "$100 now or $1100 in 10 years"; "$9 now or $100 in 10 years"; and "$40 immediately or $1000 in 10 years". A composite score was created by summing these five items. A higher score on this composite indicated a preference to wait for the larger amount of money. The mean composite score on this task was 17.82 ($SD = 6.27$).

Otherside thinking. Two issues that the college-student participants were likely to have strong opinions about were used for this task. One issue pertained to raising the cost of the tuition (Toplak & Stanovich, 2003), and the second issue pertained to banning cars on campus. In the first and opinion-gathering part of the task, participants were asked to give their opinions on each of the issues using a six-point scale ranging from disagree strongly to agree strongly. The issues read as follows:

(1) Consider the following issue: The real cost of a university education is $15000/year. Students are currently paying approximately $6000/year in tuition. The difference is paid for by the taxpayer. Indicate to what extent you agree or disagree with the following statement: University students should pay for the full cost of their university education.
(2) Consider the following issue: There is considerable debate about whether students should drive cars or take public transit to get to the York University campus. Indicate to what extent you agree or disagree with the following statement: Cars should be banned from campus.

The reason-generation part of the task was administered at a later point in the battery. This time the description of each issue was repeated, following instructions to "Think through the following issue carefully and feel free to take your time. Please type your arguments both for and against this

position. Try to write as much as you can, and remember to try and give reasons both for *and* reasons against your position. Please type your responses and label each argument (i.e., 1., 2., 3., etc.)."

A scoring scheme was developed for distinguishing between the conceptually unique reasons participants listed as being for and against the raising tuition and banning cars issues. The listed reasons were independently scored by two trained coders who used a previously developed coding scheme for classifying the statements. The inter-rater agreement between the two coders was moderate to high. Using Pearson's *r,* the inter-rater agreement between the two coders was .90 and .78, respectively, for the number of reasons to raise and not raise tuition. The inter-rater agreement was .96 and .92, respectively, for the number of reasons to ban and not ban cars from campus. In cases where the two coder's scores were not in identical agreement, a third coder independently resolved the discrepancy and determined the score.

Each participant's number of otherside reasons was derived based on his/her prior opinion. If a participant endorsed *not raising tuition* in his/her prior opinion, the number of reasons generated *in favour of raising tuition* was scored as the number of otherside reasons, and vice-versa. The total number of each participant's unique other side reasons for the raising tuition ($M = 1.36$, $SD = 1.03$) and banning cars ($M = 1.85$, $SD = 1.34$) issues were then derived based on their responses during the opinion-gathering part of the task.

Framing. In three problems, participants chose between riskless and risky alternatives that had identical expected values, under both a gain-framing and a loss-framing condition. One problem was an adaptation of Tversky and Kahneman's (1981) famous disease problem, and the two remaining problems were based on items used by Bruine de Bruin, Parker, and Fischhoff (2007; high school dropout and investment problems). Each problem was separated in the battery by other rational thinking tasks. The following scale was used to indicate preferences: (1) strongly favour option 1; (2) favour option 1; (3) slightly favour option 1; (4) slightly favour option 2; (5) favour option 2; (6) strongly favour option 2.

Problems were scored by subtracting their positive frame ratings from their corresponding negative frame ratings. Negative difference scores indicated a framing effect in the expected direction and represented a violation of the principle of descriptive invariance. Difference scores of 0 indicated the absence of a framing effect. Significant framing effects in the expected direction were found for each of the three framing problems (M differences $= -.44$ to $-.56$); $t(159) = -4.25$ to -4.32, $p < .001$. A composite framing score was formed by summing the three difference scores and reflecting the resulting values with the result that higher score indicated more resistance to framing. The mean composite score across the three items was 2.28 ($SD = 2.2$),

$t(159) = 13.33$, $p < .001$. For the purposes of the composite score, difference scores of framing effects in the unexpected direction were set to 0.

Bias blind spot. Participants read short descriptions of four specific cognitive biases: framing, base rate neglect, myside bias, and cell phone hazard, and rated the likelihood that they or a fellow students would have the bias (see the Appendix of West, Meserve, & Stanovich, 2012, for the full protocols). The framing effect item provides an example:

> Psychologists have shown that people tend to evaluate statements, arguments, or policies differently depending on the choice of words. This means that people's opinions of the very same policy or decision or product can be manipulated by slight changes in wording that don't change the meaning. For example, a food item labelled "98% fat free" is judged more attractive than one labelled "contains 2% fat". When people's opinions are manipulated based on a rewording that does not change the meaning, this is termed a framing effect.
>
> (a) To what extent do you believe that you are likely to be susceptible to framing effects?
> (b) To what extent do you believe that the average York University student is likely to be susceptible to framing effects?

Responses to the likelihood questions were given on a six-point Likert-type scale anchored at 1 (Not at all likely) and 6 (Very highly likely). Participants rated other students as more likely to commit the bias than themselves (all ps $< .01$). A composite bias blind spot score was derived by summing the self/other difference scores for the four items ($M = 1.74$, $SD = 2.18$), $t(159) = 10.07$, $p < .001$; Cronbach's alpha was .57.

RESULTS

Table 1 displays the intercorrelations among the seven CRT items as well as CRT3, CRT4, and CRT7; 44 of the 45 correlations are statistically significant. At the bottom of the table are the means and standard deviations of the variables. Performance on CRT3 was quite low, averaging less than a half an item correct. Performance on CRT4 averaged almost one item correct. More importantly, the four new items displayed a moderately high correlation of .58 with the three original items. There were indications that the four new items, although valid CRT items, were somewhat less perfect indicators than the original three items. The median correlation among the three CRT3 items was .40, whereas the median correlation among the items of CRT4 was .21. All three CRT3 items had the intuitive (wrong) answer as the modal response. The intuitive response was given 85.6%, 75.2%, and 60.0% of the time for the bat/ball, widgits, and lily pad problems, respectively. The

TABLE 1
Correlations among the CRT variables

	CRT composites			*Individual CRT items*						
Variable	1	2	3	4	5	6	7	8	9	10
1. CRT3	1									
2. CRT4	.58	1								
3. CRT7	(.86)	(.92)	1							
4. Bat/ball	(.79)	.46	(.68)	1						
5. Widgits	(.79)	.38	(.63)	.59	1					
6. Lily pads	(.81)	.52	(.73)	.38	.40	1				
7. Barrel	.44	(.67)	(.64)	.36	.32	.37	1			
8. Marks	.46	(.63)	(.62)	.39	.27	.41	.32	1		
9. Pig	.33	(.60)	(.54)	.27	.19	.31	.16	.24	1	
10. Stocks	.28	(.64)	(.54)	.18	.20	.26	.24	.18	.12	1
Mean	.49	.98	1.47	.12	.10	.28	.23	.15	.26	.34
SD	.85	1.07	1.71	.32	.30	.45	.42	.36	.44	.48

$N = 160$. CRT3 = 3 original CRT items; CRT4 = 4 new CRT items; CRT7 = all 7 CRT items; Bat/ball = CRT item 1; Widgits = CRT item 2; Lily pads = CRT item 3; Barrel = CRT item 4; Marks = CRT item 5; Pig = CRT item 6; Stocks = CRT item 7. $r = .15$, $p < .05$; $r = .20$, $p < .01$; $r = .26$, $p < .001$ (two-tailed). Correlations in parentheses reflect part–whole relationships

intuitive response was, in fact, the modal response for each of the CRT4 problems as well, but it was a less-dominant response, being given 31.3%, 51.9%, 41.9%, and 53.1% of the time for the barrel, marks, pig, and stocks problems, respectively.

Nevertheless, most indications were that the CRT4 items could be combined with the CRT3 items. The internal consistently of the seven items, considered together, was quite substantial. All 21 inter-item correlations were positive and all but one were statistically significant. The median inter-item correlation was .27, the mean was .29, and Cronbach's alpha was 0.72. Interestingly, a six-item CRT scale without the most classic and prototypical item of all—the bat and ball item—still had a Cronbach's alpha of 0.67. In short, the four new items had a substantial correlation with the classic three and when amalgamated to form a seven-item scale, the composite scale displayed substantial internal consistency.

Frederick (2005) reported that there was a highly significant sex difference on CRT3—a difference of almost a half an item correct. We replicated his finding of significantly better performance by males on CRT3 ($M = .81$ versus .29), $t(158) = 3.94$, $p < .001$. Likewise, we found a gender difference of about the same magnitude and in the same direction on CRT4 ($M = 1.38$ versus .71), $t(158) = 4.03$, $p < .001$. The effect size of .652 for CRT4 (Cohen's d) was similar to the effect size of .637 for CRT3.

TABLE 2
Correlations of CRT7, CRT3, and CRT4 with the other variables in the study

	CRT7	*CRT3*	*CRT4*
WASI	.50	.48	.41
NFC	.31	.25	.30
AOT	.42	.29	.45
ST (reflected)	.19	.15	.19
CFC	.30	.21	.32
Belief Bias in Syllogistic Reasoning	.57	.55	.48
Selection Task	.20	.22	.15
Denominator Neglect	.42	.37	.38
Temporal Discounting	.16	.15	.14
Otherside Thinking – Tuition	.29	.21	.29
Otherside Thinking – Ban Cars	.26	.24	.23
Framing	.05	.06	.03
Bias Blind Spot	.03	–.01	.05
College Average	.25	.23	.21
Rational Thinking Composite	.56	.52	.48
Thinking Dispositions Composite	.41	.30	.42

$N = 160$. CRT7 = all 7 CRT items; CRT3 = 3 original CRT items; CRT4 = 4 new CRT items; WASI = Wechsler Abbreviated Scale of Intelligence composite; NFC = Need for Cognition; AOT = Actively Openminded Thinking; ST = Superstitious Thinking: CFC = Consideration of Future Consequences. $r = .15$, $p < .05$; $r = .20$, $p < .01$; $r = .26$, $p < .001$ (two-tailed).

Table 2 displays the correlations of CRT7, CRT3, and CRT4 with the remaining variables in the study—the measure of cognitive ability (WASI), the four thinking dispositions, the seven rational thinking tasks, and college average (as well as two composite scores to be described). Of the 48 correlations in the table, 41 were statistically significant. From the table we can see that CRT3 is a bit more strongly related to intelligence than is CRT4 (.48 vs .41), although the difference was not statistically significant using a test for difference between dependant correlations (Cohen & Cohen, 1983, p. 57), $t(157) = 1.10$, *ns*. The CRT4 displayed larger correlations than CRT3 with each of the four thinking dispositions, but only the difference between the AOT correlations (.45 vs .29) reached statistical significance using a test for difference between dependant correlations, $t(157) = 2.44$, $p < .01$.

Regarding the rational thinking tasks, two of the seven rational thinking measures (framing and the bias blind spot) failed to display correlations with any of the CRT measures (and did not correlate with cognitive ability either), but the other five did. CRT4 was positively correlated with four of the tasks (belief bias, selection task, otherside thinking, and temporal discounting).

Table 3 presents a series of simultaneous regressions that examine whether CRT3 and CRT4 explain unique or overlapping variance.

TABLE 3
Simultaneous regressions examining whether CRT3 and CRT4 explain unique or overlapping variance

	Standardised Beta	*t(157)*	*Unique Variance Explained*
Criterion variable = Need for Cognition			
CRT3	0.123	1.32	0.01
CRT4	0.223	2.39*	0.033
Overall Regression: $F(2, 157) = 8.44^{***}$			
Multiple $R = .32$			
Multiple $R^2 = .10$			
Criterion variable = AOT			
CRT3	0.044	0.51	0.001
CRT4	0.422	4.81***	0.118
Overall Regression: $F(2, 157) = 19.85^{***}$			
Multiple $R = .45$			
Multiple $R^2 = .20$			
Criterion variable = Resistance to Superstitious Thinking			
CRT3	0.054	0.56	0.002
CRT4	0.159	1.65	0.017
Overall Regression: $F(2, 157) = 3.10^{*}$			
Multiple $R = .20$			
Multiple $R^2 = .04$			
Criterion variable = Consideration of Future Consequences			
CRT3	0.415	0.49	0.001
CRT4	0.29	3.12**	0.056
Overall Regression: $F(2, 157) = 8.87^{***}$			
Multiple $R = .32$			
Multiple $R^2 = .10$			
Criterion variable = Belief Bias			
CRT3	0.415	5.20***	0.114
CRT4	0.235	2.94**	0.036
Overall Regression: $F(2, 157) = 40.47^{***}$			
Multiple $R = .58$			
Multiple $R^2 = .34$			
Criterion variable = Selection Task			
CRT3	0.204	2.13*	0.027
CRT4	0.031	0.32	0.001
Overall Regression: $F(2, 157) = 4.11^{*}$			
Multiple $R = .22$			
Multiple $R^2 = .05$			
Criterion variable = Denominator Neglect			
CRT3	0.231	2.60**	0.035
CRT4	0.243	2.74**	0.039
Overall Regression: $F(2, 157) = 17.07^{***}$			
Multiple $R = .42$			
Multiple $R^2 = .18$			

(*continued*)

TABLE 3
(*Continued*)

	Standardised Beta	*t(157)*	*Unique Variance Explained*
Criterion variable = Temporal Discounting			
CRT3	0.098	1.01	0.006
CRT4	0.082	0.85	0.005
Overall Regression: $F(2, 157) = 2.08$			
Multiple $R = .17$			
Multiple $R^2 = .03$			
Criterion variable = Otherside – Tuition			
CRT3	0.06	0.64	0.002
CRT4	0.257	2.74^{**}	0.044
Overall Regression: $F(2, 157) = 7.54^{***}$			
Multiple $R = .30$			
Multiple $R^2 = .09$			
Criterion variable = Otherside – Ban Cars			
CRT3	0.155	1.64	0.016
CRT4	0.139	1.47	0.013
Overall Regression: $F(2, 157) = 5.77^{**}$			
Multiple $R = .26$			
Multiple $R^2 = .07$			
Criterion variable = College Average			
CRT3	0.168	1.77	0.019
CRT4	0.113	1.19	0.008
Overall Regression: $F(2, 157) = 5.27^{**}$			
Multiple $R = .24$			
Multiple $R^2 = .06$			

$N = 160$. CRT7 = all seven CRT items; CRT3 = three original CRT items; CRT4 = four new items; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Regarding the thinking dispositions, CRT4 was the most potent predictor in all four cases, and explained significant unique variance in three of four cases. Regarding the five rational thinking tasks that did show a relationship with the CRT, the results were more mixed but favoured CRT3 as the more potent predictor. On the belief bias task both CRT3 and CRT4 predicted unique variance once the other was accounted for, but CRT3 had a larger beta weight (and, redundantly of course, more unique variance explained). On the selection task CRT3 explained significant unique variance once CRT4 was in the regression equation, but the converse was not true. On the denominator neglect task both CRT3 and CRT4 predicted unique variance once the other was accounted for (3.5% and 3.9%, respectively), and both were about equally strong unique predictors. Neither variable was a very good predictor of temporal discounting—CRT3 and CRT4 explained similarly small amounts of unique variance in that task. CRT4 was a more

potent unique predictor of otherside arguments in the tuition task, but both CRT3 and CRT4 were roughly equal modest predictors of otherside arguments in the ban cars task and of the college average.

Table 4 presents some regression analyses run on composite variables. A rational thinking composite score was formed from the five rational thinking tasks that correlated with CRT. The *z*-scores of performance on each of the five were summed to form the rational thinking composite (the tuition and ban cars otherside *z*-scores were summed first, so that the otherside task was

TABLE 4
Additional simultaneous regression analyses

	Standardised Beta	*t(156 or 157)*	*Unique Variance Explained*
Criterion variable = Rational Thinking Composite			
CRT3	0.359	4.42^{***}	0.085
CRT4	0.269	3.31^{**}	0.048
Overall Regression: $F(2, 157) = 35.87^{***}$			
Multiple $R = .56$			
Multiple $R^2 = .31$			
Criterion variable = Rational Thinking Composite			
CRT7	0.407	5.26^{***}	0.114
Cognitive Ability	0.227	3.02^{**}	0.038
Thinking Dispositions Composite	0.084	1.18	0.006
Overall Regression: $F(3, 156) = 29.01^{***}$			
Multiple $R = .60$			
Multiple $R^2 = .36$			
Criterion variable = Rational Thinking Composite			
Cognitive Ability	0.392	5.31^{***}	0.136
Thinking Dispositions Composite	0.195	2.64^{**}	0.034
Overall Regression: $F(2, 157) = 25.39^{***}$			
Multiple $R = .49$			
Multiple $R^2 = .24$			
Criterion variable = CRT7			
Cognitive Ability	0.405	5.75^{***}	0.145
Thinking Dispositions Composite	0.272	3.87^{***}	0.065
Overall Regression: $F(2, 157) = 35.93^{***}$			
Multiple $R = .56$			
Multiple $R^2 = .31$			
Criterion variable = College Average			
CRT7	–0.035	–0.42	0.001
Cognitive Ability	0.272	3.35^{***}	0.054
Thinking Dispositions Composite	0.358	4.63^{***}	0.103
Overall Regression: $F(3, 156) = 17.37^{***}$			
Multiple $R = .50$			
Multiple $R^2 = .25$			

$N = 160$. CRT7 = all seven CRT items; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

represented by just one *z*-score, as were the other variables). The four thinking dispositions were standardised and the four *z*-scores added together to form a thinking dispositions composite score.

The first regression in Table 4 indicates that CRT3 and CRT4 were both independent predictors of the rational thinking composite score (8.5% and 4.8% of the variance, respectively). Consistent with the analyses in Table 3, the four new items, CRT4, contribute to increased predictive accuracy for these five rational thinking tasks, taken as a set. The second regression analysis in Table 4 regresses the rational thinking composite on the three main predictor variables: CRT7, cognitive ability, and the thinking dispositions composite. This analysis addresses the question investigated in Toplak et al. (2011)—whether the CRT predicts rational thinking merely because of its association with cognitive ability and thinking dispositions. In this simultaneous regression CRT7 is the dominant unique predictor, explaining 11.4% unique variance, followed by the cognitive ability (WASI) which explained a statistically significant 3.8% unique variance. The thinking dispositions composite did not predict unique variance. The third regression in Table 4 removes CRT7 from the equation and demonstrates that thinking dispositions are a significant independent predictor (3.4% unique variance) when only cognitive ability is partialled out.

In the fourth regression in Table 4 we examined which individual difference variables are predictors of CRT7 by regressing the latter on cognitive ability and the thinking dispositions. Both variables were significant independent predictors of CRT7, with cognitive ability explaining 14.5% unique variance and the thinking dispositions composite explaining 6.5% unique variance. The final analysis in Table 4 regresses college average on CRT7, cognitive ability, and the thinking dispositions composite. In the simultaneous regression only thinking dispositions and cognitive ability predicted unique variance, with the former being the stronger independent predictor (10.3% unique variance versus 5.4% unique variance).

DISCUSSION

In this study we were remarkably successful in establishing a parallel four-item version of Frederick's (2005) Cognitive Reflection Test. The four-item measure, CRT4, could be used in research as an alternative to the classic three-item measure. The need for an alternative is clear because of the increasing exposure that the three classic items are getting—particularly the bat-and-ball problem, which has appeared in books and magazines and is a common classroom demonstration. Alternatively, a seven-item version could be used, providing a more comprehensive test and one that would submerge (that is, attenuate) the effect of any contamination from the bat-and-ball item.

The reason for our optimistic conclusion regarding CRT4 and CRT7 derives from many different findings of our study. First of all, regarding CRT7, Table 1 indicates that there is substantial commonality among the seven items. The median correlation among the seven items was .27 and Cronbach's alpha was a substantial .72. Table 2 indicates that, for several of the rational thinking tasks, CRT7 was a better predictor than either CRT3 or CRT4. This was particularly true for the belief bias syllogisms and denominator neglect. Table 4 indicates that CRT7 is a substantial independent predictor of a group of rational thinking tasks—specifically, the tasks in the rational thinking composite score. The second analysis in that table indicates that CRT7 was a more potent predictor of the rational thinking composite than was either cognitive ability or thinking dispositions. After the latter two variables were entered into the equation, CRT7 still explained substantial unique variance (11.4%).

Our results demonstrated that CRT4 does, in some cases, contribute incrementally to the predictive power of CRT7. There may be a variety of reasons for this. First, the longer measure will of course be more reliable. Also, the CRT3 is known to be a difficult test and scores on it are very low even among elite populations (Frederick, 2005). Among non-elite samples floor effects might be a problem. Our sample, and a sample at the University of Toledo (see Frederick, 2005), answered only one-half of one item correct. The CRT4, on the other hand, is at least somewhat easier than the CRT3. The mean probability of answering an item correct on the CRT3 is .17, whereas the mean probability of answering an item correct on the CRT4 is .24.

As Table 3 indicates, particularly regarding the dispositions, CRT4 was a more potent predictor than CRT3. In three of four cases CRT4 accounted for significant unique variance in the thinking disposition once that the variance explained by CRT3 had been partialled out. Likewise, CRT4 was a significant unique predictor of performance on the belief bias syllogisms even after the variance attributable to CRT3 had been partialled out. The same was true for the denominator neglect task and for one of the otherside thinking measures.

Our data show that a researcher who wished to substitute CRT4 for CRT3 (perhaps due to the familiarity issues discussed above) would be amply justified, given our results. The two forms had a substantial .58 correlation. CRT3 and CRT4 had similar correlations with the rational thinking composite score (.52 and .48, respectively, see Table 2). CRT4 actually had a higher correlation with the thinking dispositions composite score than did CRT3 (.42 versus .30).

Neither CRT measure predicted performance on the framing task or in the bias blind spot task. However, measures of cognitive ability and thinking dispositions also failed to correlate with these two tasks. This lack of

association of individual difference variables with performance on the bias blind spot task is not surprising because we have previously found it to be fairly independent of cognitive ability (West, Meserve, & Stanovich, 2012). Framing performance also tends to be independent of cognitive ability when assessed in a between-participants context (Stanovich & West, 2008), but sometimes shows relationships with cognitive ability in within-participants designs like this one (Stanovich & West, 1998b). Perhaps it is because the two versions were widely separated in our battery that framing failed to correlate with either the CRT or cognitive ability.

Finally, it should be noted that the CRT had its largest correlations with the two tasks (belief bias and denominator neglect) that, according to our taxonomy (Stanovich, 2011; Stanovich, West, & Toplak, 2011), are the two tasks in the current battery that most closely represent the rational thinking category of miserly processing.

Consistent with the results of Toplak et al. (2011), in this study we demonstrated that an expanded CRT was a substantial unique predictor of rational thinking performance independent not only of cognitive ability, but also of a fairly comprehensive set of thinking dispositions. This predictive power derives, we speculate, because the CRT is a strong indicator of the miserly processing that, in dual process theory, is the source of much non-normative responding. On this view, the CRT could be interpreted as an actual *measure* of rational thought, rather than as 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 towards 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. The latter measures computational power that is *available* to the individual, but not necessarily the depth of processing that is *typically* used in most situations. Intelligence tests do not assess the tendency towards miserly processing in the way that the CRT does. Instead, in the CRT, the tendency to accept Type 1 responses is measured in a real performance context where people are searching for an accurate solution.

However, classifying the CRT as a rational thought indicator does not mean that it will correlate equally with every rational thinking task because rational thinking, in our framework, is multifarious (Stanovich, 2011; Stanovich et al., 2011). The CRT carries variance due to algorithmic cognitive capacity (the WASI in the present study) and thus will have variable correlations with other rational thinking tasks because the latter have very variable correlations with cognitive capacity (Stanovich & West, 2008). Correlations with the CRT will also tend be higher with tasks where non-normative responding is due to miserly processing. However, not all failures of rational thinking are of this type (Stanovich & West, 2008). Less-rational

responding can also be due to the fact that people have not acquired the proper declarative knowledge in domains such as scientific thinking, probabilistic reasoning, and financial and economic literacy.

REFERENCES

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.

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.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 3–11). Hillsdale, NJ: Erlbaum.

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.

Dominowski, R. (1994). Insight and instructions. *Annual Conference, British Psychological Society, Cognitive Psychology Section* (pp. 1–3). Cambridge, UK: New Hall.

Dominowski, R. L. (1995). Content effects in Wason's selection task. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp. 41–65). Hove, England: Erlbaum.

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. (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., & Stanovich, K. E. (2013). Dual process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241.

Fernbach, P. M., Sloman, S. A., Louis, R. S., & Shube, J. N. (2013). Explanation fiends and foes: How mechanistic detail determines understanding and preference. *Journal of Consumer Research*, *39*, 1115–1131.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42.

Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking & Reasoning*, *15*, 355–376.

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

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, *50*, 109–135.

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. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.

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.

Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition*, *38*, 667–676.

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

Margolis, H. (1987). *Patterns, thinking, and cognition*. Chicago: University of Chicago Press.

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*, 11–17.

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*, 353–373.

Moritz, B. B., Hill, A. V., & Donohue, K. (2013). Individual differences in the newsvendor problem: Behavior and cognitive reflection. *Journal of Operations Management*, *31*, 72–85.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*, 147–152.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*, 335–346.

Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in god. *Journal of Experimental Psychology: General*, *141*, 423–428.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

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, NJ: Erlbaum.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.

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

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York: Oxford University Press.

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 & Reasoning*, *4*, 193–230.

Stanovich, K. E., & West, R. F. (1998b). Individual differences in framing and conjunction effects. *Thinking & 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. (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.

Stanovich, K. E., & West, R. F. (2008). 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* (pp. 784–826). New York: 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.

Sunstein, C. R. (2013). *Simpler: The future of government*. New York: Simon & Schuster.

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, NJ: Erlbaum.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *94*, 197–209.

Toplak, M. E. & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology*, *17*, 851–860.

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

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.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmonsworth, England: Penguin.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: The Psychological Corporation.

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, *103*, 506–519.