On the Relative Independence of Thinking Biases and Cognitive Ability

Keith E. Stanovich
University of Toronto

Richard F. West
James Madison University

In 7 different studies, the authors observed that a large number of thinking biases are uncorrelated with cognitive ability. These thinking biases include some of the most classic and well-studied biases in the heuristics and biases literature, including the conjunction effect, framing effects, anchoring effects, outcome bias, base-rate neglect, "less is more" effects, affect biases, omission bias, myside bias, sunk-cost effect, and certainty effects that violate the axioms of expected utility theory. In a further experiment, the authors nonetheless showed that cognitive ability does correlate with the tendency to avoid some rational thinking biases, specifically the tendency to display denominator neglect, probability matching rather than maximizing, belief bias, and matching bias on the 4-card selection task. The authors present a framework for predicting when cognitive ability will and will not correlate with a rational thinking tendency.

Keywords: thinking biases, heuristics and biases, intelligence, cognitive ability

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 standard 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. In this article, we attempt to explore this assumption by examining whether measures of cognitive ability ignore important aspects of thinking itself.

A prime candidate for a cognitive domain not assessed by tests of intelligence would seem to be the domain of thinking biases. The study of heuristics and biases has been an active topic in cognitive psychology for over 3 decades now (Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 1973, 1996, 2000; Tversky & Kahneman, 1974, 1983, 1986). Many of these thinking biases relate to issues of rationality rather than to cognitive capacity (see Over, 2004; Samuels & Stich, 2004; Shafir & LeBoeuf, 2002; Stanovich, 1999, 2004), and individual differences in their operation remain unassessed on IQ tests.

However, there may be reasons for expecting a relationship between cognitive ability and individual differences in the operation of thinking biases even if the latter are not directly assessed on intelligence tests. This follows from the fact that theorizing in the heuristics and biases literature has emphasized dual-process models of cognition (Evans, 2003, 2006, 2007; Kahneman, 2000, 2003; Kahneman & Frederick, 2002, 2005; Sloman, 1996; Stanovich, 1999, 2004). Such models embody the assumption that thinking biases are universal properties of a heuristic system that operates autonomously (called System 1 in dual-process theory; see Kahneman & Frederick, 2002; Stanovich, 1999). Nonetheless, although the presence of thinking biases might be universal, their ability to result in nonnormative choices varies from individual to individual because heuristic responses are sometimes overridden by a nonautonomous analytic system of thought (called System 2 in dual-process theory). The computational power needed to override a heuristically primed response and/or the ability to recognize the need to override might be related to intelligence, thus creating a (negative) relationship between biased responding and cognitive ability even though thinking biases are not directly assessed on IQ tests.

Some research has indicated that cognitive ability is modestly related to performance on several tasks from the heuristics and biases literature. Stanovich and West (1997, 1998c, 1998d, 1999, 2000; see also Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Sá, West, & Stanovich, 1999; Toplak & Stanovich, 2002) found correlations with cognitive ability to be roughly (in absolute magnitude) .35–.45 for belief bias in syllogistic reasoning, in the range of .25–.35 for various probabilistic reasoning tasks, in the range of .20–.25 for various covariation detection and hypothesis

Keith E. Stanovich, Department of Human Development and Applied Psychology, University of Toronto, Ontario, Canada; Richard F. West, Department of Graduate Psychology, James Madison University.

This research was supported by grants from the Social Sciences and Humanities Research Council of Canada and the Canada Research Chairs program to Keith E. Stanovich.

Correspondence concerning this article should be addressed to Keith E. Stanovich, Department of Human Development and Applied Psychology, University of Toronto, 252 Bloor Street West, Toronto, Ontario M5S 1V6, Canada. E-mail: kstanovich@oise.utoronto.ca

Journal of Personality and Social Psychology, 2008, Vol. 94, No. 4, 672–695

0022-3514/08/$12.00 DOI: 10.1037/0022-3514.94.4.672

testing tasks, .25–.35 on informal reasoning tasks, .15–.20 with outcome bias measured within subjects, .20–.40 with performance in the four-card selection task, .10–.20 with performance in various disjunctive reasoning tasks, .15–.25 with hindsight bias, .25–.30 with denominator neglect, and .05–.20 with various indices of Bayesian reasoning. All correlations were in the expected direction, and all were from studies with 100+ participants. Other investigators have found relationships of a similar effect size between cognitive ability and a variety of tasks in the heuristics and biases literature (Bruine de Bruin, Parker, & Fischhoff, 2007; DeShon, Smith, Chan, & Schmitt, 1998; Handley, Capon, Beveridge, Dennis, & Evans, 2004; Klaczynski & Lavallee, 2005; Newstead, Handley, Harley, Wright, & Farrelly, 2004; Parker & Fischhoff, 2005; Perkins & Ritchhart, 2004; Valentine, 1975).

In a commentary on this research on individual differences, Kahneman (2000) pointed out that the correlations observed may well have been inflated because most of the relevant studies used within-subjects designs, which contain cues signalling the necessity of heuristic system override (Bartels, 2006; Fischhoff, Slovic, & Lichtenstein, 1979; Kahneman & Frederick, 2002; Kahneman & Tversky, 1982a; Shafir, 1998). He argued that between-subjects tests of the coherence of responses represent a much stricter criterion and perhaps a more appropriate one because "much of life resembles a between-subjects experiment" (Kahneman, 2000, p. 682).

That the mental factors operative in within-subjects designs might be different from those operative in between-subjects designs suggests that the individual difference factors associated with biased processing in the two different paradigms might also vary. LeBoeuf and Shafir (2003) have produced some data indicating that biases that are assessed within-subjects display different relationships with individual difference variables than biases assessed between-subjects. They found that various framing effects were associated with the need for cognition thinking disposition (see Cacioppo, Petty, Feinstein, & Jarvis, 1996) when evaluated on a within-subjects basis but were independent of need for cognition when framing was assessed between subjects.

We know little about the relation between cognitive ability and the tendency to make coherent judgments in between-subjects situations. Thus, Kahneman's (2000; see also Kahneman & Frederick, 2002) conjecture that these less transparent designs would reduce the observed relationships between cognitive ability and the judgmental biases mentioned above remains virtually untested. In the present experiments, we examined a variety of effects from the heuristics and biases literature to see if cognitive ability was associated with these biases as they are displayed in between-subjects paradigms. In the first experiment, we examined some biases and effects that are among the oldest in the literature (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1973, 1974, 1983): base-rate neglect, framing effects, conjunction effects, anchoring biases, and outcome bias.

Experiment 1: Classic Heuristics and Biases Effects

Method

Participants

The participants were 434 undergraduate students (102 men and 332 women) recruited through an introductory psychology subject pool at a medium-sized state university in the United States. Their mean age was 19.0 years (*SD* = 1.7). The participants were randomly assigned to Form A (216 participants) and Form B (218 participants).

Experimental Tasks

Base-rate problem. Kahneman and Tversky's (1973) much-studied lawyer/engineer problem was employed as a probe of the degree of base-rate usage. The two versions were identical except that the base rates given for the engineers and lawyers were switched as a between-subjects variable (30 engineers and 70 lawyers in Form A and 70 engineers and 30 lawyers in Form B). Both groups responded by estimating the probability that the focal individual was one of the engineers.

Framing problem (Asian disease). This item was based on Tversky and Kahneman's (1981) famous disease problem. The positive (gain) and negative (loss) framing of the problem was a between-subjects manipulation (Form A was the gain frame and Form B was the loss frame). Both groups of participants chose a response on the following scale: "I strongly favor program A" (1); "I favor program A" (2); "I slightly favor program A" (3); "I slightly favor program B" (4); "I favor program B" (5); "I strongly favor program B" (6). Higher scored responses represented more risk seeking.

Conjunction problem. This problem was based on Tversky and Kahneman's (1983) much-studied Linda problem. Participants read the following: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations."

Participants then used a 6-point scale (1 = *extremely improbable*, 2 = *very improbable*, 3 = *somewhat probable*, 4 = *moderately probable*, 5 = *very probable*, 6 = *extremely probable*) to indicate the relative probability of three statements that described Linda. The first two statements were identical for the two groups of participants as follows: 1) It is ______ that Linda is a teacher in an elementary school; 2) It is ______ that Linda works in a bookstore and takes Yoga classes. Each group then read one of two statements that differed in whether they did or did not contain a conjunction of two descriptions. Participants getting Form A read the following: 3) It is ______ that Linda is a bank teller. Participants getting Form B read the following: 3) It is ______ that Linda is a bank teller and is active in the feminist movement.

Outcome bias. Our measure of outcome bias derived from a problem investigated by Baron and Hershey (1988). Participants receiving Form A read the positive outcome version involving a 55-year-old man who had a heart condition and whose operation succeeded. Participants evaluated the physician's decision to go ahead with the operation (1 = *incorrect, a very bad decision*; 2 = *incorrect, all things considered*; 3 = *incorrect, but not unreasonable*; 4 = *the decision and its opposite are equally good*; 5 = *correct, but the opposite would be reasonable too*; 6 = *correct, all things considered*; 7 = *clearly correct, an excellent decision*).

Participants receiving Form B (negative outcome) evaluated a medical decision that was designed to be objectively better than the first: 2% chance of death rather than 8%; 10-year increase in life expectancy versus 5-year increase, etc. However, it had an unfortunate negative outcome—death of the patient.

Anchoring and adjustment problems. The two problems used here were adapted from an anchoring and adjustment problem in Tversky and Kahneman (1974) and one used by Epley and Gilovich (2004). Prior to making an estimation of a particular value, participants answered a question containing a small or large anchor value. The Form A version is given below, with the Form B value in brackets:

1. Do you think there are more or less than 65 [12] African countries in the United Nations? (a. more; b. less); How many African countries do you think are in the United Nations? _________

2. Is the tallest redwood tree in the world more that 85 [1000] feet tall? (a. more; b. less); How tall do you think the tallest redwood tree in the world is? _________

Cognitive Ability Measure

Students were asked to indicate their verbal, mathematical, and total SAT scores on the demographics form. The mean reported verbal SAT score of the students was 577 ($SD = 68$), the mean reported mathematical SAT score was 572 ($SD = 69$), and the mean total SAT score was 1149 ($SD = 110$). The institution-wide averages for this university in 2006 were 565, 575, and 1140, respectively. Several studies have indicated that the correlation between self-reported SATs and verified SAT scores is in the range of .80 to .92 (Cassady, 2001; Kuncel, Crede, & Thomas, 2005; Noftle & Robins, 2007) as is the correlation between self-reported and verified grade-point average (Higgins, Peterson, Pihl, & Lee, 2007). An indication of the validity of the self-reported scores is that they correlated with a third variable to the same extent as verified scores. Stanovich and West (1998c) found that the correlation between a vocabulary test and self-reported SAT total scores (.49) was quite similar to the .51 correlation between the vocabulary test and verified total SAT scores in a previous investigation that used the same vocabulary measure (West & Stanovich, 1991). These indications of validity are perhaps consistent with the fact that participation in these experiments represents a low-stakes, anonymous situation in which participants have little reason to misrepresent their SAT scores (in contrast to a more high-stakes situation where a job or some other benefit may be on the line).

The total SAT score was used as an index of cognitive ability in the analyses reported here because it loads highly on psychometric *g* (Frey & Detterman, 2004; Unsworth & Engle, 2007). For the purposes of some of the analyses described below, the 206 students with SAT scores below the median (1150) were assigned to the low-SAT group, and the 228 remaining students were assigned to the high-SAT group. Parallel analyses that are fully continuous and that did not involve partitioning the sample are also reported.[1]

One caveat concerning the associations in these studies relates to the restriction of range in our sample. Certainly, it is true that individuals with average and above average cognitive ability are overrepresented in samples composed entirely of university students. Nevertheless, the actual range in cognitive ability found among college students in the United States is quite large. In the past 30 years, the percentage of 25- to 29-years-olds in the United States who have attended college has increased by 50%. By 2002, 58% of these young adults had completed at least 1 or more years of college, and 29% had received at least a bachelor's degree (U.S. Department of Health and Human Services, 2003). However, the restriction of range in cognitive ability is somewhat greater in our sample because our participants attended a moderately selective state university. The SAT total means of our samples are roughly .60 of a standard deviation above the national mean of 1021 (College Board, 2006). The standard deviation of the distribution of scores in our sample is roughly .55–.70 of the standard deviation in the nationally representative sample.

Results

Table 1 displays, for each of the experimental tasks, the mean response as a function of Form (A vs. B) and cognitive ability group (low vs. high SAT). The table also contains, for each of the experimental tasks, an analysis that examines whether the magnitude of the effect or bias demonstrated by the task was moderated by cognitive ability. This was done by examining, in an analysis of variance (ANOVA) context, whether the effect of form interacted with SAT group.

The first analysis indicated that there was a significant base-rate effect on the engineer/lawyer problem, $F(1, 430) = 35.93$, $MSE = 472.2$, $p < .001$. The mean probability that Jack was one of the engineers was lower in the 30% base-rate condition (60.8% for the entire sample) than the mean probability that Jack was one of the engineers in the 70% base-rate condition (73.4% for the entire sample). However, the effect of base rate failed to interact with cognitive ability, as the Form × SAT Group interaction was not significant, $F(1, 430) = 1.30$, $MSE = 472.2$. To the extent that there is any hint of an interaction in the means, it is in the opposite direction from the expected finding. The low-SAT group was slightly more sensitive to base rates than was the high-SAT group.

The next task displayed in Table 1 is the disease framing task, and it is clear that both groups displayed the expected framing effect on this problem—the loss frame (Form B) resulted in a greater preference for the risky option. The main effect of frame type (form) was highly significant, $F(1, 430) = 50.98$, $MSE = 1.37$, $p < .001$. However, the effect of frame type failed to interact with cognitive ability, as the Form × SAT Group interaction was not significant, $F(1, 430) = 1.48$, $MSE = 1.37$. To the extent there

[1] Separate analyses involving the verbal SAT (SAT–V) and the mathematical SAT (SAT–M) score were conducted in parallel to all of the analyses in this article that employ the SAT total score. Across the studies, there were few differences between the analyses of the total score and those employing either of the components, but those that did occur are listed in this footnote. In Experiment 1, the interaction involving the conjunction problem occurred for SAT–V but not for SAT–M, and the interaction involving outcome bias occurred for SAT–M but not for SAT–V. In Experiment 5, omission bias was more frequent (28.1% vs. 18.5%) in the group lower in SAT–V, $\chi^2(1, N = 458) = 5.88$, $p < .05$, but there was no difference associated with SAT–M. In probabilistic reasoning Problem 1 of Experiment 8, both SAT–M and SAT–V were correlated with the maximizing response, but the correlation involving SAT-M was higher. In probabilistic reasoning Problem 1, the point biserial correlation between responding normatively (MAX) versus nonnormatively (MATCH or OTHER) and SAT–M was .269, whereas the corresponding correlation with SAT–V was .151. In probabilistic reasoning Problem 2, the point biserial correlations between responding normatively (MAX) versus nonnormatively (MATCH or OTHER) and SAT–M and SAT–V were similar (.183 and .222, respectively).

Table 1
ANOVA (Form × SAT) and Mean Scores as a Function of Form (A vs. B) and SAT in Experiment 1; Results From a Parallel Regression Analysis Are Also Indicated

Source	$F(1, 430)$	η_p^2	Cognitive ability	Form A M (SD)	Form B M (SD)
Base-rate problem (engineer/lawyer problem)				30 engineers	70 engineers
Form	35.93***	.077			
SAT	4.47*	.010	Low SAT	57.4 (25.6)	72.3 (18.2)
Form × SAT	1.30	.003	High SAT	64.2 (26.5)	74.3 (14.6)
Form × SAT interaction in regression: $F = 1.43$, R^2 change for interaction = .003					
Framing problem (Asian disease)				Gain frame	Loss frame
Form	50.98***	.106			
SAT	4.81*	.011	Low SAT	3.00 (1.24)	3.67 (1.13)
Form × SAT	1.48	.003	High SAT	3.11 (1.16)	4.05 (1.16)
Form × SAT interaction in regression: $F = 0.08$, R^2 change for interaction = .001					
Conjunction problem (Linda problem)				Bank teller	Feminist bank teller
Form	120.5***	.219			
SAT	0.24	.001	Low SAT	2.53 (1.03)	3.46 (1.13)
Form × SAT	4.66*	.011	High SAT	2.36 (0.98)	3.73 (1.19)
Form × SAT interaction in regression: $F = 4.15^*$, R^2 change for interaction = .008					
Outcome bias				Positive outcome	Negative outcome
Form	20.50***	.045			
SAT	10.09**	.023	Low SAT	5.79 (1.07)	5.12 (1.26)
Form × SAT	3.88*	.009	High SAT	5.91 (0.87)	5.65 (1.04)
Form × SAT interaction in regression: $F = 4.34^*$, R^2 change for interaction = .009					
Anchoring (African countries)				Large anchor	Small anchor
Form	219.1***	.338			
SAT	1.37	.003	Low SAT	45.2 (26.4)	14.4 (14.4)
Form × SAT	2.53	.006	High SAT	40.0 (22.7)	15.2 (11.0)
Form × SAT interaction in regression: $F = 3.82$, R^2 change for interaction = .006					
Anchoring (redwoods)				Small anchor	Large anchor
Form	461.0***	.520			
SAT	0.09	.000	Low SAT	126.1 (88.4)	977.4 (580.4)
Form × SAT	0.05	.000	High SAT	128.6 (77.3)	998.5 (580.6)
Form × SAT interaction in regression: $F = 0.19$, R^2 change for interaction = .001					

Note. *df* on redwoods = 426. ANOVA = analysis of variance.
$^* p < .05$. $^{**} p < .01$. $^{***} p < .001$.

is any hint of an interaction in the means, it is again in the opposite direction from the expected finding. The high-SAT group displayed a slightly larger framing effect.

The next task displayed in Table 1 is the conjunction problem (Linda problem). The means of both groups displayed the expected conjunction fallacy—Linda was judged more probably a feminist bank teller than a bank teller. The main effect of form was highly significant, $F(1, 430) = 120.5$, $MSE = 1.18$, $p < .001$. There was a significant Form × Cognitive Ability interaction, $F(1, 430) = 4.66$, $MSE = 1.18$, $p < .05$, although the size of the effect was small ($\eta_p^2 = .011$). Additionally, the interaction was in the opposite direction from the expected finding—the high-SAT group was more susceptible to the conjunction fallacy.

The next task displayed in Table 1 is the outcome bias problem. The means of both groups displayed the expected outcome bias—the decision with the positive outcome was rated as a better decision than the decision with the negative outcome, despite the fact that the latter was objectively better. The main effect of outcome (form) was highly significant, $F(1, 430) = 20.50$, $MSE = 1.13$, $p < .001$. There was a significant Form × Cognitive Ability interaction, $F(1, 430) = 3.88$, $MSE = 1.13$, $p < .05$, although the size of the effect was small ($\eta_p^2 = .009$). This interaction was in the expected direction—the low-SAT group displayed more outcome bias.

The remaining two analyses in Table 1 concern the two anchoring and adjustment problems. In the African countries item, it is clear that both groups displayed the expected anchoring effect—the large anchor resulted in higher estimates of the number of African countries in the United Nations ($M = 42.6$ for the entire sample) than did the small anchor ($M = 14.9$ for the entire sample). The number of African countries in the United Nations is actually 53. The main effect of anchor magnitude (form) was highly significant, $F(1, 430) = 219.1$, $MSE = 381.4$, $p < .001$. However, the effect of anchor magnitude failed to interact with cognitive ability, as the Form × SAT Group interaction was not significant, $F(1, 430) = 2.53$, $MSE = 381.4$.

Results were similar for the redwoods problem. Both groups displayed the expected anchoring effect—the large anchor resulted in higher estimates of the height of the tallest redwood (mean of 989.0 ft for the entire sample) than did the small anchor (mean of 127.4 ft for the entire sample). The tallest redwood tree is actually 370 ft in height. The main effect of anchor magnitude (form) was highly significant, $F(1, 426) = 461.0$, $MSE = 171{,}862$, $p < .001$. However, the effect of anchor magnitude failed to interact with cognitive ability, as the Form × SAT Group interaction was not significant, $F(1, 426) = 0.05$, $MSE = 171{,}862$.

Thus, across all of the ANOVAs, only two of the six interactions between form and cognitive ability were statistically significant, and only one of these (that for outcome bias) was in the direction of reduced judgmental bias on the part of the group higher in cognitive ability. Of course, some information is lost—and power reduced—by dichotomizing on the cognitive ability variable (SAT). Thus, Table 1 also presents the results of a continuous analysis for each task in which the significance of the Form × Cognitive Ability interaction was tested in a regression analysis in which SAT was used as a continuous variable rather than as a dichotomous variable. Immediately below the ANOVA in Table 1 is presented the F ratio for the test of the interaction: the Form × SAT cross product when entered third in the equation predicting item response after form and SAT. Along with the F ratio is presented the R^2 change for the interaction. These and all subsequent regression analyses were conducted with centered predictors.

The regression analyses converged completely with the results from the ANOVAs. Only two of the six interaction terms reached statistical significance, and only one of those (that for outcome bias) was in the expected direction—the degree of outcome bias was larger for the group lower on the SAT. The significant interaction term for the Linda problem indicated a data pattern in the opposite direction—the conjunction fallacy was displayed to a greater extent by the group with higher SAT scores.

In summary, Experiment 1 produced very little evidence indicating that cognitive ability was related to judgmental biases when the latter were assessed in between-subjects designs. In six comparisons involving five different classic effects from the heuristics and biases literature, only one comparison provided any indication that cognitive ability attenuated a judgmental bias. Even in that case (outcome bias), the effect was extremely modest. The variance associated with the interaction effect was less than 1% in both the ANOVA and in the regression analysis.

Experiment 2: "Less is More" Effects

It has been proven through several formal analyses that if people's preferences follow certain logical patterns (the so-called axioms of choice), then they are behaving as if they are maximizing utility (Dawes, 1998; Edwards, 1954; Jeffrey, 1983; Luce & Raiffa, 1957; Savage, 1954; von Neumann & Morgenstern, 1944). Much empirical work has gone into determining whether humans adhere to the axioms of choice (transitivity, independence, reduction of compound lotteries, etc.). However, a between-subjects comparison makes it possible to test an even simpler requirement of rational choice—that people can be framed into preferring more to less (e.g., preferring $6 to $5). For example, Slovic, Finucane, Peters, and MacGregor (2002) found that people rated a gamble with 7/36 chance to win $9 and 29/36 to lose 5¢ more favorably than a gamble with 7/36 chance to win $9 and 29/36 chance to win nothing. Indeed, they reported that the latter gamble was even rated less desirable than a gamble having a 7/36 chance to win $9 and 29/36 to lose 25¢. Presumably, in a between-subjects design, the representation of the numerically small loss highlights the magnitude and desirability of the $9 to be won.

Likewise, the phenomenon of proportion dominance can result in people preferring less to more in a between-subjects design. Slovic et al. (2002) reported a study in which people rated a safety measure that would save 150 lives less favorably than a safety measure that would save 98% of 150 lives at risk (see also Bartels, 2006; Slovic & Peters, 2006). Their explanation of this "less is more" effect exhibited in a between-subjects design is that saving 150 lives is more diffusely good than saving 98% of some target figure because the 98% is more evaluable (see Bartels, 2006; Hsee, Loewenstein, Blount, & Bazerman, 1999; Hsee & Zhang, 2004)—it is close to the upper bound on a percentage scale. In the present experiment, we examined whether any of these "less is more" effects attenuate as the cognitive ability of the participant sample increases.

Method

Participants

The participants were 361 undergraduate students (149 men and 212 women) recruited through an introductory psychology subject pool at a medium-sized state university. Their mean age was 19.2 years ($SD = 1.5$). None had participated in Experiment 1. The participants were randomly assigned to Form A (122 participants), Form B (119 participants), or Form C (120 participants).

The mean reported verbal SAT score of the students was 586 ($SD = 68$), the mean reported mathematical SAT score was 596 ($SD = 66$), and the mean total SAT score was 1182 ($SD = 107$). For the purposes of some of the analyses described below, the 183 students with SAT scores at or below the median (1190) were assigned to the low-SAT group, and the 178 remaining students were assigned to the high-SAT group.

Experimental Tasks

Gamble. Participants receiving Form A evaluated the following proposition: "I would find a game that had a 7/36 chance of winning $9 and a 29/36 chance of winning nothing extremely attractive." Participants responded on a 6-point scale: *strongly agree* (6), *moderately agree* (5), *slightly agree* (4), *slightly disagree* (3), *moderately disagree* (2), and *strongly disagree* (1). Participants receiving Form B evaluated the following proposition: "I would find a game that had a 7/36 chance of winning $9 and a 29/36 chance of losing 5¢ extremely attractive." Participants receiving Form C evaluated the following proposition: "I would find a game that had a 7/36 chance of winning $9 and a 29/36 chance of losing 25¢ extremely attractive."

Proportion dominance Problem 1. Participants receiving Form A evaluated the following version:

> Imagine that highway safety experts have determined that a substantial number of people are at risk of dying in a type of automobile fire. A requirement that every car have a built-in fire extinguisher (estimated cost, $300) would save the 150 people who would otherwise die every year in this type of automobile fire. Rate the following statement about yourself: I would be extremely supportive of this requirement.

Participants responded on a 6-point scale ranging from *strongly agree* (6) to *strongly disagree* (1). Participants receiving Form B evaluated a version in which the text read "98% of the 150 people" in place of "the 150 people." Participants receiving Form C evaluated a version in which the text read "95% of the 150 people" in place of "the 150 people."

Proportion dominance Problem 2. Participants receiving Form A evaluated the following version:

> You have recently graduated from the University, obtained a good job, and are buying a new car. A newly designed seatbelt has just become available that would save the lives of the 500 drivers a year who are involved in a type of head-on-collision. (Approximately half of these fatalities involve drivers who were not at fault.) The newly designed seatbelt is not yet standard on most car models. However, it is available as a $500 option for the car model that you are ordering. How likely is it that you would order your new car with this optional seatbelt?

Participants responded on a 6-point scale ranging from *very likely* to *very unlikely*. Participants receiving Form B evaluated a version in which the text read "98% of the 500 drivers" in place of "the 500 drivers." Participants receiving Form C evaluated a version in which the text read "95% of the 500 drivers" in place of "the 500 drivers."

Results

Table 2 displays, for each of the three "less is more" problems, the mean response as a function of Form (A, B, C) and cognitive ability group (low vs. high). The Table also contains, for each of the three problems, an analysis that examines whether the magnitude of the "less is more" effect was moderated by cognitive ability. This was done by examining whether the effect of form interacted with SAT group—first in an ANOVA context and then in a fully continuous regression analysis.

The first analysis indicated that there was a significant "less is more" effect on the gamble problem. The higher expected value gamble with no possibility of loss was rated as a less appealing

Table 2
ANOVA (Condition × SAT) and Mean Scores as a Function of Condition and SAT in Experiment 2; Results From a Parallel Regression Analysis Are Also Indicated

Source	F	η_p^2	Cognitive ability	Form A	Form B	Form C
				M (SD)	M (SD)	M (SD)
Gamble				7/36 win $9 29/36 nothing	7/36 win $9 29/36 lose 5¢	7/36 win $9 29/36 lose 25¢
Condition	12.54^{***}	.066				
SAT	0.08	.000	Low SAT	2.17 (1.27)	3.73 (1.62)	3.28 (1.73)
Condition × SAT	5.91^{**}	.032	High SAT	2.70 (1.48)	2.88 (1.93)	3.46 (1.66)
Form × SAT interaction in regression: $F = 6.53^{**}$, R^2 change for interaction = .033						
Proportion dominance #1				150 saved	98% of 150 saved	95% of 150 saved
Condition	5.02^{**}	.028				
SAT	0.51	.001	Low SAT	4.44 (1.31)	4.62 (1.16)	4.61 (1.31)
Condition × SAT	1.78	.010	High SAT	4.20 (1.31)	4.82 (1.10)	4.92 (1.06)
Form × SAT interaction in regression: $F = 2.40$, R^2 change for interaction = .013						
Proportion dominance #2				500 saved	98% of 500 saved	95% of 500 saved
Condition	3.66^{*}	.020				
SAT	0.01	.000	Low SAT	4.33 (1.42)	4.41 (1.43)	4.57 (1.37)
Condition × SAT	1.27	.007	High SAT	4.00 (1.70)	4.61 (1.30)	4.73 (1.33)
Form × SAT interaction in regression: $F = 1.28$, R^2 change for interaction = .007						

Note. dfs = 1,355 on SAT and 2,355 on the other two effects.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

gamble than either of the lower expected value gambles that contained the possibility of loss, $F(2, 355) = 12.54$, $MSE = 2.62$, $p < .001$. There was a significant Condition × Cognitive Ability interaction, $F(2, 355) = 5.91$, $MSE = 2.62$, $p < .01$, although the effect was modest ($\eta_p^2 = .032$). This interaction was in the expected direction—the low-SAT group displayed more of a "less is more" effect. The high-SAT group rated the lose-5¢ condition similarly to the lose-nothing condition, whereas the low-SAT group rated the latter as considerably less attractive. However, the impressively rational performance of the high-SAT group in the lose-5¢ condition was somewhat impeached by their performance in the lose-25¢ condition, which they rated as considerably more attractive than the lose-nothing condition. Thus, both groups proved susceptible to a "less is more" effect, but the high-SAT simply proved somewhat less susceptible.

The next analysis indicated that there was also a significant "less is more" effect on proportion dominance Problem 1. The proposal for the fire extinguisher that saved 95% of 150 lives and the one that saved 98% of 150 lives were rated higher than the proposal for the fire extinguisher that saved 150 lives, $F(2, 355) = 5.02$, $MSE = 1.46$, $p < .01$. There was no significant Condition × Cognitive Ability interaction, $F(2, 355) = 1.78$, $MSE = 1.46$, *ns*. To the extent that there is any hint of an interaction in the means, it is in the opposite direction from the expected finding. The high-SAT group displayed a slightly larger "less is more" effect on this problem.

The last analysis indicates that there was also a significant "less is more" effect on proportion dominance Problem 2. The seatbelt that saved 95% of 500 lives and the one that saved 98% of 500 lives were rated more highly than the seatbelt that saved 500 lives, $F(2, 355) = 3.66$, $MSE = 2.04$, $p < .05$. There was no significant Condition × Cognitive Ability interaction, $F(2, 355) = 1.27$, $MSE = 2.04$, *ns*. As with proportion dominance Problem 1, any hint of an interaction in the means is in the opposite direction from the expected finding. The high-SAT group displayed a slightly larger "less is more" effect on this problem.

Also presented in Table 2 are the results of regression analyses in which SAT was used as a continuous variable rather than as a dichotomous variable. Immediately below the ANOVA in Table 2 is presented the *F* ratio for the test of the interaction: the Form × SAT cross product when entered third in the equation predicting item response after form and SAT. Along with the *F* ratio is presented the R^2 change for the interaction. The regression analyses converged completely with the results from the ANOVAs. Only the interaction in the gamble problem was statistically significant.

This experiment produced very little evidence that cognitive ability attenuates "less is more" effects in between-subjects designs. In two items that display a proportion dominance effect, there was no statistically significant tendency for higher SAT participants to show less proportion dominance. On the gamble problem, there was a tendency for the "less is more" effect to be smaller among high-SAT participants. The size of the interaction was small, however ($\eta_p^2 = .032$). Also, high-SAT participants did display a robust "less is more" effect when the comparison involved the 25¢-loss condition—ironically, the one with the lowest expected value.

Experiment 3: Honoring Sunk Costs and Absolute Versus Relative Savings

In this experiment, we examined the nonnormative economic behavior of honoring sunk costs—the tendency to persist in a negative expected value activity because a significant investment has already been made (Arkes & Ayton, 1999; Arkes & Blumer, 1985). Additionally, we examined the economically inefficient tendency to pursue the maximum relative savings rather than the maximum absolute savings (Thaler, 1980; Tversky & Kahneman, 1981).

Method

Participants

The 729 participants in this experiment (249 men and 480 women) were recruited as in the previous two experiments. None had participated in Experiments 1 or 2. The mean reported total SAT score was 1167 ($SD = 102$). For the purposes of the analyses described below, the 364 students with SAT scores below the median (1180) were assigned to the low-SAT group, and the 365 remaining students were assigned to the high-SAT group.

Experimental Tasks

Sunk-cost fallacy. The participants were randomly assigned to the no-sunk-cost condition and the sunk-cost condition. The no-sunk-cost group read the following problem:

> Imagine that Video Connection rents videos for $1.50. Low Priced Videos, a competitor, rents videos for just $2.00 each. Although the Video Connection store is ten minutes away by car, the Low Priced Videos is only about 1/2 block from your apartment. Assuming that you only rent from these two stores, how many of your next 20 videos rentals would be from the closer but more expensive Low Priced Videos?

Participants then responded on a 10-point scale: *0–1* (1), *2–3* (2), *4–5* (3), *6–7* (4), *8–9* (5), *10–11* (6), *11–12* (7), *13–14* (8), *15–16* (9), and *17–20* (10). Responses to this item establish how likely participants are to choose convenience over a discount when they have not already paid for the privilege of the discount. The sunk-cost group read a similar scenario but one in which the choice was between convenience and a discount when the cost of the privilege of the discount had already been sunk:

> Imagine that you just paid $50 for a Video Connection discount card that allows you to rent videos for 50% off the regular price of $3.00. Soon after you purchased the Video Connection discount card, Low Priced Videos, a competitor, opened a new store that rents videos for just $2.00 each. Although the Video Connection store is ten minutes away by car, the new Low Priced Videos is only about 1/2 block from your apartment. Assuming that you only rent from these two stores, how many of your next 20 videos rentals would be from the closer but more expensive Low Priced Videos?

Participants responded on the same 10-point scale.

Absolute versus relative savings. This task was based on comparison problems introduced by Thaler (1980; see also Tversky & Kahneman, 1981). Participants received either a large percentage of savings (Form 1) or a small percentage of savings (Form 2) framing form of the problem. The Form 1 version is given below:

> Imagine that you go to purchase a calculator for $30. The calculator salesperson informs you that the calculator you wish to buy is on sale

for $20 at the other branch of the store which is ten minutes away by car. Would you drive to the other store? Option A: Yes, Option B: No

The Form 2 version is given below:

Imagine that you go to purchase a jacket for $250. The jacket salesperson informs you that the jacket you wish to buy is on sale for $240 at the other branch of the store which is ten minutes away by car. Would you drive to the other store? Option A: Yes, Option B: No

Previous research has found more participants willing to make the trip to save $10 for the calculator than for the jacket, thus violating the standard analysis of consumer behavior, which views the two versions as equivalent choices between traveling and gaining $10 versus the status quo (Thaler, 1980; Tversky & Kahneman, 1981).

Results

The sunk-cost paradigm used in this experiment was successful in creating a sunk-cost effect. The mean score of 6.71 ($SD = 2.9$) in the no-sunk-cost condition (corresponding to roughly 11 videos) was significantly higher than the mean score of 4.61 ($SD = 3.2$) in the sunk-cost condition (corresponding to roughly 7.5 videos), $t(727) = 9.33$, $p < .001$, Cohen's $d = 0.692$. Both cognitive ability groups displayed sunk-cost effects of roughly equal magnitude. For the high-SAT group, the mean in the no-sunk-cost condition was 6.90 and the mean in the sunk-cost condition was 5.08, whereas for the low-SAT group, the mean in the no-sunk-cost condition was 6.50 and the mean in the sunk-cost condition was 4.19. A 2 (cognitive ability) × 2 (condition) ANOVA indicated a significant main effect of cognitive ability, $F(1, 725) = 8.40$, $MSE = 9.13$, $p < .01$, and a significant main effect of condition, $F(1, 725) = 84.9$, $MSE = 9.13$, $p < .001$. There was a slight tendency for the low-SAT participants to show a larger sunk-cost effect, but the Cognitive Ability × Condition interaction did not attain statistical significance, $F(1, 725) = 1.21$, $MSE = 9.13$. The interaction was also tested in a regression analyses in which SAT was treated as a continuous variable rather than as a dichotomous variable. The Form × SAT cross product, when entered third in the equation, was not significant, $F(1, 725) = 0.32$.

The sunk-cost effect thus represents another cognitive bias that is not strongly attenuated by cognitive ability. However, this is true only when it is assessed in a between-subjects context. Using a similar sunk-cost problem, Stanovich and West (1999) did find an association with cognitive ability when participants responded in a within-subjects design.

Results on the percentage versus absolute savings problem were consistent with previous research. More participants were willing to drive 10 min when the resulting $10 savings constituted a substantially larger percentage of the cost of the purchase. For the high-SAT group, 88.3% were willing to drive to save $10 on the $30 calculator, but only 55.9% would drive to save $10 on the $250 jacket, $\chi^2(1, N = 365) = 47.14$, $p < .001$. These values for the low-SAT group were, respectively, 86.9% and 62.0%, $\chi^2(1, N = 364) = 30.11$, $p < .001$. Of particular interest in the present study, a logistic regression analysis indicated that the interaction between SAT group and problem form failed to reach significance, $\beta = .38$, $\chi^2(1, N = 729) = 1.01$. The nonnormative preference for relative over absolute savings was not moderated by cognitive ability.

Experiment 4: Nonseparability of Risk and Benefit Judgments

Slovic et al. (2002; Slovic & Peters, 2006) have suggested that judgments about the risks and benefits of various activities and technologies derive not from separable knowledge sources relevant to risk and benefit but instead derive from a common source—affect. Evidence for this conjecture derives from the finding that ratings of risk and reward are negatively correlated (Finucane, Alhakami, Slovic, & Johnson, 2000; Slovic & Peters, 2006), both across activities within participants and across participants within activities. When something is rated as having high benefits, it tends to be seen as having low risk, and when something is rated as having high risk, it is seen as having low benefits. Finucane et al. (2000) argued that such a finding is nonnormative because the risk/benefit relationship is most likely positive in a natural ecology. Their argument is that of the four cells in a high/low partitioning of risk and benefit, one of the cells—that for activities of high risk and low benefit—must be vastly underpopulated. This is because activities of this type are usually not adopted, and they are often proscribed by authorities even when they are. If the high-risk, low-benefit quadrant is underpopulated, then the overall risk/benefit of activities in the actual world must be positively correlated. In the present study, we investigated whether the nonnormative tendency to view risk and reward as negatively correlated is attenuated in individuals high in cognitive ability.

Method

The 458 participants in this experiment (112 men and 346 women) were recruited as in the previous experiments, and their mean age was 18.8 years ($SD = 1.8$). None of them had participated in previous experiments. The mean reported verbal SAT score of the students was 565 ($SD = 74$), the mean reported mathematical SAT score was 570 ($SD = 78$), and the mean total SAT score was 1135 ($SD = 127$). The 231 students with SAT scores at or below the median (1140) were assigned to the low-SAT group, and the 227 remaining students were assigned to the high-SAT group. Participants were asked to rate the benefits and risks of the following technologies/products: bicycles, alcoholic beverages, chemical plants, pesticides. Participants were first asked to rate the technology/product on a 7-point scale ranging from *not at all beneficial* (1) to *very beneficial* (7). They then rated the relative risk of the technology/product on a 7-point scale ranging from *not at all risky* (1) to *very risky* (7).

Results

Finucane et al. (2000) used three methods to assess the relationship between judgments of benefit and risk. First, they examined the mean risk/benefit ratings given to 23 different activities/technologies and then calculated the correlation across the 23 items. Second, they calculated the correlation, across the 23 items, of the ratings of each individual participant. Finally, Finucane et al. (2000) examined, within each activity, whether those who rated the activity as having large benefits perceived it as having low risk and vice versa. All three of these methods yielded a preponderance of negative correlations. Because we examined only four different activities/technologies, we were restricted to the third method of calculating the correlation between risk and benefit.

Table 3 presents the mean benefit and risk rating for each of the four activities/technologies. Not surprisingly, bicycles were rated of highest benefit and lowest risk. Alcoholic beverages, in contrast were rated relatively low in benefit and high in risk. Most relevant, however, are the correlations between benefit and risk within each of the four items. These correlations are presented for both the high-SAT and low-SAT group. Replicating the work of Finucane et al. (2000), all eight of the correlations were negative and seven of the eight were statistically significant. For each of the four activities/technologies, people who saw the activity as very beneficial saw it as lower in risk, and those who saw it as lower in benefit tended to see it as somewhat higher in risk.

The nonnormative tendency to view risk and reward as negatively correlated was not attenuated by high cognitive ability within the range that we studied. In fact, each of the four negative correlations were higher in the high-SAT group than in the low-SAT group. However, none of the differences between the correlations were significant when a Fisher *z* test (values of 1.51, 0.12, 0.45, and 1.12, respectively) was used.

Experiment 5: Omission Bias

Omission bias is the tendency to avoid actions that carry some risk but that would prevent a larger risk (Baron, 1998; Baron & Ritov, 2004; Bazerman, Baron, & Shonk, 2001). People often do not realize that by failing to act they often subject themselves and others to greater risk. In Experiment 5, we examined whether the extent of omission bias was attenuated by cognitive ability.

Method

The participants were 236 of the participants who completed the tasks of Experiment 4. Participants read the following problem:

> Imagine that there will be a deadly flu going around your area next winter. Your doctor says that you have a 10% chance (10 out of 100) of dying from this flu. However, a new flu vaccine has been developed and tested. If taken, the vaccine prevents you from catching the deadly flu. However, there is one serious risk involved with taking this vaccine. The vaccine is made from a somewhat weaker type of flu virus, and there is a 5% (5 out of 100) risk of the vaccine causing you to die from the weaker type of flu. Imagine that this vaccine is completely covered by health insurance. If you had to decide now, which would you choose?
>
> 1. I would definitely not take the vaccine. I would thus accept the 10% chance of dying from this flu. [scored as 1]
> 2. I would probably not take the vaccine. I would thus accept the 10% chance of dying from this flu. [scored as 2]
> 3. I would probably take the vaccine. I would thus accept the 5% chance of dying from the weaker flu in the vaccine. [scored as 3]
> 4. I would definitely take the vaccine. I would thus accept the 5% chance of dying from the weaker flu in the vaccine. [scored as 4]

Results

Explicit omission bias (the tendency to avoid the treatment—that is choosing either Alternative 1 or 2 on the scale) was a minority phenomenon among both cognitive ability groups. However, it was equally common in both groups. In the high-SAT group, 25.6% of the participants displayed some degree of omission bias, whereas 30.3% of the low-SAT group displayed some degree of omission bias, a difference that was not statistically significant, $\chi^2(1, N = 236) = 0.62$. As a convergent, fully continuous analysis we correlated the response on the scale with the SAT total score and observed a nonsignificant correlation of .057. Thus, neither analysis indicated that there was statistically reliable evidence that omission bias was associated with cognitive ability in range studied in this investigation.

Experiment 6: Reference Dependent Preferences—WTA/WTP and the Certainty Effect

The theory of reference-dependent preferences (Kahneman & Tversky, 1979, 1984; Tversky & Kahneman, 1991) predicts a large difference in willingness-to-accept (WTA) and willingness-to-pay (WTP) valuations. In this experiment, we examined individual differences in WTA/WTP discrepancies and also another effect predicted by prospect theory (Kahneman & Tversky, 1979)—the certainty effect. Prospect theory predicts that people overweight probability differences that make an outcome certain over similar probability differences that do not (a violation of the standard assumptions of utility theory).

Method

Thaler problem. The 164 participants (70 men and 94 women) completing this problem in this experiment were recruited as in the previous experiments. The 86 students with SAT scores below the median (1160) were assigned to the low-SAT group, and the 78 remaining students were assigned to the high-SAT group.

Table 3
Means and Correlation Between Benefit and Risk for the High-SAT and Low-SAT Groups in Experiment 4

Variable	Benefit rating *M* (*SD*)	Risk rating *M* (*SD*)	Correlation: High-SAT group	Correlation: Low-SAT group	Fisher *z*
Bicycles	5.49 (1.25)	3.16 (1.32)	$-.184^{**}$	$-.039$	1.51
Alcoholic beverages	3.03 (1.31)	5.22 (1.25)	$-.310^{***}$	$-.300^{***}$	0.12
Chemical plants	3.85 (1.26)	4.86 (1.34)	$-.261^{***}$	$-.220^{***}$	0.45
Pesticides	4.15 (1.25)	4.61 (1.34)	$-.274^{***}$	$-.170^{**}$	1.12

$^{**} p < .01$. $^{***} p < .001$.

Participants were randomly assigned to one of three conditions—WTP certainty, WTA certainty, and WTP noncertainty. WTP-certainty and WTA-certainty conditions were based on a problem discussed by Thaler (1980). The WTP-certainty group read the following problem:

> Imagine that when you went to the movies last week, you were inadvertently exposed to a rare and fatal virus. The possibility of actually contracting the disease is 1 in 1,000, but once you have the illness there is no known cure. On the other hand, you can, readily and now, be given an injection that stops the development of the illness. Unfortunately, these injections are only available in very small quantities and are sold to the highest bidder. What is the highest price you would be prepared to pay for such an injection? [You can get a long-term, low-interest loan if needed.]: $10 [scored as 1]; $100 [scored as 2]; $1,000 [scored as 3]; $10,000 [scored as 4]; $50,000 [scored as 5]; $100,000 [scored as 6]; $250,000 [scored as 7]; $500,000 [scored as 8]; $1,000,000 [scored as 9]; $5,000,000 or more [scored as 10; logarithmic scoring schemes resulted in essentially identical results].

The WTA-certainty group read the following problem and answered on the same scale:

> Imagine that a group of research scientists in the School of Medicine are running a laboratory experiment on a vaccine for a rare and fatal virus. The possibility of actually contracting the disease from the vaccine is 1 in 1,000, but once you have the disease there is no known cure. The scientists are seeking volunteers to test the vaccine on. What is the lowest amount that you would have to be paid before you would take part in this experiment?

The WTP-noncertainty group read the following problem and answered on the same scale:

> Imagine that when you went to the movies last week, you were inadvertently exposed to a rare and fatal virus. The possibility of actually contracting the disease is 4 in 1,000, but once you have the illness there is no known cure. On the other hand, you can, readily and now, be given an injection that reduces the possibility of contracting the disease to 3 in 1,000. Unfortunately, these injections are only available in very small quantities and are sold to the highest bidder. What is the highest price you would be prepared to pay for such an injection? [You can get a long-term, low-interest loan if needed.]

The Allais problem. The participants in this experiment were the same as those who participated in Experiment 3. The Allais problem was modeled on Allais's (1953) famous two-choice problem wherein participants sometimes violate the so-called independence axiom of utility theory—that if the outcome in some state of the world is the same across options, then that state of the world should be ignored. Approximately one half of the participants received Form 1 of the problem:

> Choose between:
> A: One million dollars for sure
> B: .89 probability of one million dollars
> .10 probability of five million dollars
> .01 probability of nothing

Approximately one half of the participants received Form 2 of the problem:

> Choose between:
> C: .11 probability of one million dollars
> .89 probability of nothing
> D: .10 probability of five million dollars
> .90 probability of nothing

The typical finding is that most people select Option A in Form 1 and Option D in Form 2, but these choices violate the independence axiom (the outcome of the .89 slice of probability is the same in A and B as it is in C and D). Although many theorists have discussed why individuals finding Option D attractive might nonetheless be drawn to Option A in the first problem (Bell, 1982; Loomes & Sugden, 1982; Maher, 1993; Schick, 1987; Slovic & Tversky, 1974), relatively little is known about the influence that cognitive ability might have on preferences on this problem.

Results

There was a significant WTA/WTP difference displayed in Experiment 6. The mean score of 7.49 ($SD = 2.89$) in the WTA-certainty condition was significantly higher than the mean score of 5.76 ($SD = 2.68$) in the WTP-certainty condition, $t(105) = 3.21$, $p < .01$, Cohen's $d = 0.627$). Participants would require between $250,000 and $500,000 to take on the 1 in 1,000 risk of the disease but would pay only $50,000 to $100,000 to have the 1 in 1,000 probability of the disease removed. There was also a significant certainty effect displayed in Experiment 6. The mean score of 5.76 ($SD = 2.68$) in the WTP-certainty condition was significantly higher than the mean score of 3.87 ($SD = 2.64$) in the WTP-noncertainty condition, $t(106) = 3.69$, $p < .001$, Cohen's $d = 0.717$). Participants would pay $50,000 to $100,000 to have the probability of the disease entirely removed but only between $1,000 and $10,000 to have the likelihood of the disease reduced by the same probability.

As indicated in Table 4, both cognitive ability groups displayed WTA/WTP discrepancies and certainty effects of roughly equal magnitude. A 2 (cognitive ability) × 3 (condition) ANOVA indicated that there was a significant main effect of condition, $F(2, 152) = 20.61$, $MSE = 7.44$, $p < .001$, but no significant main effect of cognitive ability, $F(1, 152) = 3.56$, $MSE = 7.44$, $.05 < p < .10$. It is important that the Cognitive Ability × Condition interaction did not attain statistical significance, $F(2, 152) = 0.41$, $MSE = 7.44$. Neither the WTA/WTP discrepancy nor the certain effect was attenuated by cognitive ability.

The same was true of the responses to the Allais problems. Consistent with previous research on this problem, the majority of participants chose Option A for Form 1, whereas only a minority of participants selected the analogous Option C for Form 2. For the high-SAT group, the percentages for Options A and C were, respectively, 55.3% and 28.0%, $\chi^2(1, N = 365) = 28.13$, $p < .001$. These values for the low-SAT group were, respectively, 74.2% and 33.7%, $\chi^2(1, N = 364) = 60.07$, $p < .001$. Of particular interest in the present study, a logistic regression analysis indicated that the interaction between SAT group and Allais problem form failed to reach significance, $\beta = 0.57$, $\chi^2(1, N = 729) = 3.21$, $p = .073$.

Experiment 7: Myside Bias

People display myside bias when they generate evidence, test hypotheses, and evaluate policies in a manner biased toward their own opinions (Baron, 1991; Evans, 2002; Greenhoot, Semb, Colombo, & Schreiber, 2004; Johnson-Laird, 2006; Klaczynski & Lavallee, 2005; Kuhn, 2001; Perkins, 1985, 2002; Sá, Kelley, Ho, &

Table 4
ANOVA (Condition × SAT) and Mean Scores as a Function of Condition and SAT in Experiment 6; Results From a Parallel Regression Analysis Are Also Indicated

Source	F	η_p^2	Cognitive ability	WTP certainty M	WTP certainty (SD)	WTP noncertainty M	WTP noncertainty (SD)	WTA certainty M	WTA certainty (SD)
Condition	20.61***	.213							
SAT	3.56	.023	Low SAT	6.46	(2.93)	4.36	(2.40)	7.69	(3.08)
Condition × SAT	0.41	.005	High SAT	5.11	(2.30)	3.69	(2.85)	7.25	(2.69)
Form × SAT interaction in regression: $F = 0.63$, R^2 change for interaction = .006									

Note. *df*s = 1,152 on SAT and 2,152 for the other two effects.
*** $p < .001$.

Stanovich, 2005; Stanovich & West, 2007; Toplak & Stanovich, 2003). It is easy to disguise the intent of an experiment on myside bias by using a between-subjects manipulation, as we show in Experiment 7.

Method

The participants were the same 458 participants who completed the tasks of Experiments 4 and 5. Participants were randomly assigned to the myside (Ford Explorer) and otherside (German car) groups. The Ford Explorer group read the following problem:

> According to a comprehensive study by the U.S. Department of Transportation, Ford Explorers are 8 times more likely than a typical family car to kill occupants of another car in a crash. The Department of Transportation in Germany is considering recommending a ban on the sale of the Ford Explorer in Germany. Do you think that Germany should ban the sale of the Ford Explorer?

Participants answered on the following scale: *definitely yes* (6), *yes* (5), *probably yes* (4), *probably no* (3), *no* (2), *definitely no* (1). Participants were also asked: "Should the Ford Explorer be allowed on German streets, just like other cars?" They answered on the same scale as the previous question.

The German car group read the following problem:

> According to a comprehensive study by the U.S. Department of Transportation, a particular German car is 8 times more likely than a typical family car to kill occupants of another car in a crash. The U.S. Department of Transportation is considering recommending a ban on the sale of this German car.

Participants answered the following two questions on the scale presented above: (1) Do you think that the United States should ban the sale of this car? (2) Do you think that this car should be allowed on U.S. streets, just like other cars?

Results

The paradigm used in this experiment was successful in creating a myside bias effect. On the "ban the sale" of the car question, the mean for the Ford Explorer (myside) condition was significantly lower (3.61, $SD = 1.23$) than the mean for the German car (otherside) condition (4.34, $SD = 1.23$), $t(456) = -6.33$, $p < .001$, Cohen's $d = 0.593$. Participants were more likely to think that the German car should be banned in the U.S. than they were to think that the Ford Explorer should be banned in Germany. Correspondingly, on the "allowed on the streets like other cars" question, the mean for the Ford Explorer (myside) condition was significantly higher (3.82, $SD = 1.19$) than the mean for the German car (otherside) condition (2.86, $SD = 1.20$), $t(456)$, = 8.59, $p < .001$, Cohen's $d = 0.805$. Participants were more likely to think that the Ford Explorer should be allowed on German streets like other cars than they were to think that the German car should be allowed on U.S. streets like other cars.

As Table 5 indicates, both cognitive ability groups displayed

Table 5
ANOVA (Form × SAT) and Mean Scores as a Function of Condition and SAT in Experiment 7

Source	$F(1, 454)$	η_p^2	Cognitive ability	German car M	German car (SD)	Ford Explorer M	Ford Explorer (SD)
Ban car?							
Condition	39.89***	.081					
SAT	0.81	.002	Low SAT	4.42	(1.24)	3.63	(1.27)
Form × SAT	0.28	.001	High SAT	4.26	(1.21)	3.59	(1.20)
Allow car on streets like other cars?							
Condition	73.99***	.140					
SAT	2.78	.006	Low SAT	2.78	(1.20)	3.71	(1.24)
Form × SAT	0.10	.000	High SAT	2.93	(1.20)	3.93	(1.13)

*** $p < .001$.

myside bias effects of roughly equal magnitude. For the ban-the-car question, a 2 (cognitive ability) × 2 (condition) ANOVA indicated a significant main effect of condition, $F(1, 454) = 39.89$, $MSE = 1.51$, $p < .001$, but no main effect of cognitive ability, $F(1, 454) = 0.81$, $MSE = 1.51$, *ns*. The Cognitive Ability × Condition interaction did not attain statistical significance, $F(1, 454) = 0.28$, $MSE = 1.51$. A regression analysis in which cognitive ability was treated as a continuous variable confirmed the lack of interaction, $F(1, 454) = 0.52$.

Correspondingly, for the "allowed on the streets like other cars" question, a 2 (cognitive ability) × 2 (condition) ANOVA indicated a significant main effect of condition, $F(1, 454) = 73.99$, $MSE = 1.42$, $p < .001$, but no main effect of cognitive ability, $F(1, 454) = 2.78$, $MSE = 1.42$, $.05 < p < .10$. The Cognitive Ability × Condition interaction did not attain statistical significance, $F(1, 454) = 0.10$, $MSE = 1.42$. A regression analysis in which cognitive ability was treated as a continuous variable confirmed the lack of interaction, $F(1, 454) = 0.13$.

In summary, the bias toward the rights of the Ford Explorer in Germany and against the rights of the German car in the United States was not attenuated by cognitive ability in the range studied in this sample.

Experiment 8: Heuristics and Biases That Do Associate With Cognitive Ability

In the previous seven experiments, we have shown that a wide variety of cognitive biases are dissociated from cognitive ability. However, we do not mean to imply that effects from the heuristics and biases literature are invariably independent of intelligence. To the contrary, in the introduction we mentioned several biases in which there has been at least some suggestion of an association. In this experiment, we collected together several such effects, including several well-known logical reasoning and probabilistic reasoning tasks and biases.

Method and Results

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; Pacini & Epstein, 1999). The task 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 in 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 visual of each tray was presented to participants]. From which tray would you prefer to select a marble in a real situation?: ____ the small tray ____ the large tray.

The normative response is of course to select the small tray, because it has the highest probability (10% vs. 8%) of yielding a winning black marble. Nonetheless, Epstein and colleagues (see also Klaczynski, 2001; Kokis et al., 2002) have found that a significant minority of participants choose the large tray. Kokis et al. (2002) found, in a sample of children, a significant .28 correlation between cognitive ability and the tendency to choose the small tray. However, little is known about the nature of the correlation in an adult sample.

The 819 participants in this experiment (179 men and 640 women) were recruited in the same manner as in all the previous studies, but none had participated in previous experiments. Their mean reported total SAT score was 1161 ($SD = 105$). A substantial number of participants displayed denominator neglect in the marble task. Although a majority of participants (519—63.4%) picked the normatively correct small tray, 300 of the participants (36.6%) chose the large tray, thus displaying denominator neglect. The tendency to respond normatively in this task was, however, significantly associated with cognitive ability. The mean SAT score of the participants choosing the small tray (1174, $SD = 107$) was significantly higher than the mean score of those choosing the large tray (1137, $SD = 99$), $t(817) = 4.90$, $p < .001$, Cohen's $d = 0.356$. The point biserial correlation between tray choice and SAT total was .169 ($p < .001$). Unlike the effects and biases discussed in Experiments 1–7, the phenomenon of denominator neglect in this probabilistic choice paradigm was related to cognitive ability.

Probabilistic Reasoning: Probability Matching Versus Maximizing

Problem 1. Modified from West and Stanovich (2003), this problem read as follows:

> Consider the following hypothetical situation: A deck with 10 cards is randomly shuffled 10 separate times. The 10 cards are composed of 7 cards with the number "1" on the down side and 3 cards with the number "2" on the down side. Each time the 10 cards are reshuffled, your task is to predict the number on the down side of the top card. Imagine that you will receive $100 for each downside number you correctly predict, and that you want to earn as much money as possible. What would you predict after shuffle #1? (1 or 2); What would you predict after shuffle #2? (1 or 2); . . . What would you predict after shuffle #10? (1 or 2).

Of the 819 participants who completed the marble problem (denominator neglect), 440 completed this probability matching Problem 1, and 379 completed probability matching Problem 2, to be described next.

For this problem, students who predicted any combination of 7 "1" cards and 3 "2" cards were classified as using the MATCH strategy ($n = 184$, 41.8%). Students who predicted "1" for each of the 10 cards were classified as using the maximizing (MAX) strategy ($n = 105$, 23.9%). Any other combination of card guesses was classified as the OTHER strategy ($n = 151$, 34.3%). The maximizing strategy is of course normative in this paradigm, but it was used by only a minority of participants.

Probabilistic Reasoning: Probability Matching Versus Maximizing

Problem 2. Adapted from West and Stanovich (2003; see also Gal & Baron, 1996), this problem read as follows:

> Consider the following situation: A die with 4 red faces and 2 green faces will be rolled 6 times. Before each roll you will be asked to predict which color (red or green) will show up once the die is rolled.

> Which color is most likely to show up after roll #1? (1 = red or 2 = green); Which color is most likely to show up after roll #2? (1 = red or 2 = green); . . . Which color is most likely to show up after roll #6? (1 = red or 2 = green).

For this problem, students who chose any combination of 4 “red” faces and 2 “green” faces were classified as using the MATCH strategy ($n = 173$, 45.6%). Students who chose “red” for each of the six rolls were classified as using the MAX strategy ($n = 130$, 34.3%). Any other combination of color choices was classified as the OTHER strategy ($n = 76$, 20.1%).

Table 6 displays the mean SAT score for each of the response groups for probabilistic reasoning Problems 1 and 2. For Problem 1, there was a significant overall effect of group, $F(2, 437) = 16.55$, $MSE = 10{,}126$, $p < .001$, $\eta_p^2 = .070$. The mean SAT score for the MAX group was significantly higher than the mean SAT score for both the MATCH group and the OTHER group. The point biserial correlation between responding normatively (MAX) versus nonnormatively (MATCH or OTHER) and SAT total score was .262 ($p < .001$).

The results for probabilistic reasoning Problem 2 were exactly parallel. There was a significant overall effect of group, $F(2, 376) = 14.85$, $MSE = 10{,}662$, $p < .001$, $\eta_p^2 = .073$. The mean SAT score for the MAX group was significantly higher than the mean SAT score for both the MATCH group and the OTHER group. The point biserial correlation between responding normatively (MAX) versus nonnormatively (MATCH or OTHER) and SAT total score was .270 ($p < .001$). Like denominator neglect, the tendency to maximize predictive accuracy in this probabilistic reasoning task was related to cognitive ability.

Belief Bias: Syllogisms

Belief bias occurs when people have difficulty evaluating conclusions that conflict with what they think they know about the world (Evans, 2002; Evans, Barston, & Pollard, 1983; Evans, Newstead, Allen, & Pollard, 1994; Klauer, Musch, & Naumer, 2000). It is most often assessed with syllogistic reasoning tasks in which the believability of the conclusion conflicts with logical validity.

Twenty-four syllogistic reasoning problems, largely drawn from Markovits and Nantel (1989), were completed by the participants. The 436 participants for this task (127 men and 309 women) were recruited in the same manner as in all the previous studies, but none had participated in previous experiments. Their mean reported total SAT score was 1174 ($SD = 109$). Eight of the problems—the inconsistent syllogisms—were worded such that the validity judgment was in conflict with the believability of the conclusion (e.g., All flowers have petals; roses have petals; therefore, roses are flowers–which is invalid). Eight of the problems—the consistent syllogisms—were worded such that the validity judgment was congruent with the believability of the conclusion (e.g., All fish can swim; tuna are fish; therefore, tuna can swim—which is valid). Eight of the problems—representing the neutral condition—involved imaginary content (e.g., All opprobines run on electricity; Jamtops run on electricity; therefore, Jamtops are opprobines–which is invalid) and were thus neither consistent nor inconsistent. Participants were instructed to judge the validity of the conclusion assuming that all of the premises were true.

A belief bias effect was demonstrated on the syllogistic reasoning task. The mean number of items answered correctly (out of 8) was 6.80 ($SD = 1.27$) in the consistent condition, 6.73 ($SD = 1.38$) in the neutral condition, and 5.11 ($SD = 1.91$) in the inconsistent condition. Participants answered significantly more consistent items than inconsistent items, $t(435) = 18.06$, $p < .001$. Cognitive ability was significantly correlated with the magnitude of belief bias (number of consistent items correct minus number of inconsistent items correct) displayed by each participant ($r = -.28$, $p < .001$). Regarding the individual conditions separately, cognitive ability correlated significantly ($p < .001$) with the number of consistent items correct ($r = .25$), the number of neutral items correct ($r = .39$), and the number of inconsistent items correct ($r = .45$). Cognitive ability correlated more highly with the inconsistent items than with the consistent items (test for difference between dependent correlations), $t(433) = 4.07$, $p < .001$.

Belief Bias: Deductive Certainty of Modus Ponens

The 381 participants for this task (94 men and 287 women) were recruited in the same manner as in all the previous studies, but none had participated in previous experiments. Their mean reported total SAT score was 1174 ($SD = 109$). This task was adapted from a study by George (1995). It is a deductive reasoning task that assesses whether participants recognize the deductive certainty of modus ponens. Participants read four arguments. Following is one example:

> Premises: 1. If there is a postal strike, then unemployment will double. 2. There is a postal strike. Conclusion: 3. Unemployment will double.

Participants responded on the following scale: *true* (7), *probably true* (6), *somewhat true* (5), *uncertain* (4), *somewhat false* (3),

Table 6
Means (Standard Deviations) of SAT Score for Participants Using the MAX, MATCH, and OTHER Strategies on the Two Probability Matching Problems in Experiment 8

	MAX			MATCH			OTHER			
Variable	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)	ANOVA $F(2, 437)$
Probabilistic Reasoning #1	105	1207_a	(101)	184	1139_b	(105)	151	1149_b	(95)	16.55^{***}
Probabilistic Reasoning #2	130	1203_a	(107)	173	1141_b	(100)	76	1145_b	(105)	14.85^{***}

Note. Means in the same row that show different subscripts are significantly different at $p < .05$ in a Scheffé post hoc test. MAX = maximizing; ANOVA = analysis of variance.
*** $p < .001$.

probably false (2), and *false* (1). Premise 1 in the other three arguments was as follows: “If the winter is harsh, then there will be a flu epidemic,” “If a car is a Honda, then it is expensive,” “If a person eats hamburgers, then he/she will get cancer.” The score on the task was the sum of the responses to all four items. Because the normatively correct answer on each item is “true,” normatively correct responding on the task should result in a score of 28.

Belief bias was shown on at least one of the modus ponens problems by 222 participants (they were less than certain that the syllogism was true on at least one problem). In contrast, 159 participants displayed no-belief bias on this task—they answered *true* to each of the four items. The mean SAT score of the no-belief bias group (1196, $SD = 105$) was significantly higher than the mean SAT score (1164, $SD = 106$) of the participants who displayed some degree of belief bias, $t(379) = 2.91$, $p < .01$, Cohen’s $d = 0.302$. A parallel, continuous analysis was conducted in which the score on each of four items (*true* = 7, *probably true* = 6, etc.) was summed to yield a total score on the modus ponens belief bias items. The correlation between this index and the SAT total score was .19 ($p < .001$).

Informal Reasoning: Argument Evaluation Test

The same participants who completed the previous task also were given the Argument Evaluation Test (AET; for details, see Stanovich & West, 1997). The test consists of two parts. First, participants indicated their degree of agreement with a series of 23 target propositions (on topics such as gun control, taxes, university governance, crime, etc.) on a 4-point scale. Participants then evaluated arguments (which varied on an operationally defined measure of strength; see Stanovich & West, 1997) relevant to these propositions. Individual differences in participants’ reliance on objective argument quality were examined by running separate regression analyses on each participant’s responses. Each participant’s 23 argument evaluation responses were regressed simultaneously on both the 23 argument quality scores and the 23 prior opinion scores. The former beta weight was used as the primary indicator of the ability to evaluate arguments independently of their prior opinion on the issue in question.

Performance on the AET was evaluated by examining the beta weight for argument quality for each participant. This score ranged from −.506 to .768 and displayed a mean of .296 ($SD = .215$). This beta weight was greater than zero for 90.9% of the sample. The beta weight for argument quality was significantly correlated with the SAT total score ($r = .353$, $p < .001$), indicating that informal reasoning about argument quality was associated with cognitive ability.

Four-Card Selection Task

Participants who completed this task were 375 of the participants who completed the modus ponens task. Originally used by Wason (1966), the abstract version of the selection task has been studied extensively in the deductive reasoning literature (for detailed discussions of the enormous literature on the task, see Evans, Newstead, & Byrne, 1993; Evans & Over, 1996, 2004; Sperber, Cara, & Girotto, 1995; Stanovich & West, 1998a). The problem involves reasoning about the falsifiability of an “if P then Q” type of rule. We used a version with features that facilitated performance: “violation” instructions, requiring justifications for each of the choices, and using a form of the rule that discourages a biconditional interpretation (see Platt & Griggs, 1993, 1995).

The correct response is to choose the P and not-Q cards. We specifically examined the common response patterns that have been discussed in the literature: P-only, PQ, all of the cards, and PQ not-Q. All remaining response patterns were classified as Other. We also constructed the “logic index” employed by Pollard and Evans (1987). It is formed by subtracting the number of incorrect responses (not-P and Q) from the number of correct responses (P and not-Q).

The major response patterns on the four-card selection task and the mean SAT score associated with that pattern are displayed in Table 7. Because this version of the task contained several features that have been found to facilitate performance in past studies (Platt & Griggs, 1993, 1995), the solution rate (30.9%) was substantially higher than that usually obtained with the abstract selection task (often 10% or less). Cognitive ability was associated with the response given to this version of selection task. In a one-way ANOVA, there was a significant overall effect of group, $F(5, 369) = 8.22$, $MSE = 10{,}213$, $p < .001$, $\eta_p^2 = .100$. Partitioning the sample somewhat differently, we found that the mean SAT score (1214, $SD = 96$) of the group answering correctly was significantly higher than the mean SAT score (1163, $SD = 106$) of those giving an incorrect answer, $t(373) = 4.40$, $p < .001$, Cohen’s $d = 0.493$. Finally, there was a statistically significant correlation between the logic index score for the task and the SAT total score ($r = .292$, $p < .001$). Thus, all of the analyses of this task consistently pointed to an association between cognitive ability and selection task performance.

Table 7
Means (Standard Deviations) of Total SAT Score for Each of the Major Response Patterns in the Four-Card Selection Task

P, not-Q [Correct] ($n = 116$)		P ($n = 51$)		P, Q ($n = 38$)		P, Q, not-Q ($n = 39$)		All ($n = 35$)		OTHER ($n = 96$)		ANOVA
M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	$F(5, 369)$
1214_a	(96)	1179	(116)	1174	(96)	1209_a	(92)	1170	(101)	1129_b	(104)	8.22^{***}

Note. Means that show different subscripts are significantly different at $p < .05$ in a Scheffé post hoc test. ANOVA = analysis of variance.
$^{***}p < .001$.

Discussion

The studies reported here demonstrated that many biases discussed in the heuristics and biases literature are surprisingly independent of cognitive ability in the range examined in our experiments. We say surprisingly because ever since Spearman (1904) first discovered positive manifold, intelligence indicators have correlated with a plethora of cognitive abilities and thinking skills that are almost too large to enumerate (e.g., Ackerman, Kyllonen, & Richards, 1999; Carroll, 1993; Deary, 2000, 2001; Lubinski, 2000, 2004; Lubinski & Humphreys, 1997; Sternberg, 1977, 2000). Of course, the generalization of our results must be restricted to the range of cognitive ability contained in our sample. All of our studies employed university students as participants. The higher and lower SAT groupings are partitionings of the upper half and lower half of the sample—the low-SAT group is not low in an absolute sense; they simply are of lower cognitive ability relative to their counterparts in that particular study. The magnitude of the correlations involving SAT obtained in these studies is undoubtedly attenuated because of restriction of range. However, the fact that the range of the samples studied is somewhat restricted makes many of the findings (of near zero correlations between cognitive ability and many aspects of rational thought) no less startling. It is quite unexpected that, across even the range of ability in a university population, there would be so little relation between rational thought and cognitive ability.

Mindware Gaps and Override Detection

Of course, it is not true that all thinking biases are independent of cognitive ability. The right column of Table 8 lists some effects and biases from Experiment 8 and from other studies where an association was found. Kahneman (2000) offers the beginning of an explanation of why certain rational thinking tasks might show associations with cognitive ability and others may not. His argument begins with the distinction between coherence rationality and reasoning rationality. Reasoning rationality "requires an ability to reason correctly about the information currently at hand without demanding perfect consistency among beliefs that are not simultaneously evoked" (Kahneman & Frederick, 2005, p. 277). In contrast, "coherence is much stricter . . . coherence requires choices and beliefs to be immune to variations of framing and context. This is a lot to ask for, but an inability to pass between-subjects tests of coherence is indeed a significant flaw" (Kahneman, 2000, p. 682). Kahneman and Frederick (2002; see Kahneman, 2000), utilizing a dual-process framework, argued that correlations with cognitive ability will occur only in the intermediate range of difficulty. There, they argued, "intelligent people are more likely to possess the relevant logical rules and also to recognize the applicability of these rules in particular situations. In the terms of the present analysis, high-IQ respondents benefit from relatively efficient System 2 operations that enable them to overcome erroneous intuitions when adequate information is available. When a problem is too difficult for everyone, however, the correlation is likely to reverse" (Kahneman & Frederick, 2002, p. 68).

The phrase "possess the relevant logical rules and also to recognize the applicability of these rules in particular situations" suggests two conditions that have to be fulfilled for a heuristically based response to be overridden by analytic processing (Evans, 2003, 2006, 2007; Kahneman & Frederick, 2002; Stanovich, 1999). These two conditions are actually the two sources of judgmental error that Kahneman and Tversky (1982a), two-decades ago, labeled as errors of application and errors of comprehension. The latter refers to errors that occur because people do not recognize the validity of a norm that they have violated. The former occurs when a person fails to apply a rule that he or she has learned.

In the remainder of this article, we use two slightly different terms for the loci of these problems. An error of comprehension we call a mindware gap (Stanovich, 2008a, 2008b; Toplak, Liu, Macpherson, Toneatto, & Stanovich, 2007). This is because in dual-process models, an important function of the analytic system is to take early representations triggered by the heuristic system offline and to substitute better responses. Mindware, a term coined by Perkins (1995; Clark, 2001, uses it in a slightly different way

Table 8
Tasks in These Experiments and in Other Studies That Do and Do Not Show Associations With Cognitive Ability

Tasks/effects that fail to correlate with cognitive ability	Tasks/effects that correlate with cognitive ability
Noncausal base-rate usage (Experiment 1; Stanovich & West, 1998c, 1999)	Causal base-rate usage (Stanovich & West, 1998c, 1998d)
Conjunction fallacy between subjects (Experiment 1)	Outcome bias between and within subjects (Experiment 1; Stanovich & West, 1998c)
Framing between-subjects (Experiment 1, Experiment 3)	Framing within subjects (Bruine de Bruin et al. 2007; Frederick, 2005; Parker & Fischhoff, 2005; Stanovich & West, 1998b, 1999)
Anchoring effect (Experiment 1)	Denominator neglect (Experiment 8; Kokis et al., 2002)
Evaluability "less is more" effect (Experiment 2)	Probability matching (Experiment 8; West & Stanovich, 2003)
Proportion dominance effect (Experiment 2)	Hindsight bias (Stanovich & West, 1998c)
Sunk-cost effect (Experiment 3; Parker & Fischhoff, 2005)	Ignoring P(D/NH) (Stanovich & West, 1998d, 1999)
Risk/benefit confounding (Experiment 4)	Covariation detection (Sá et al., 1999; Stanovich & West, 1998c, 1998d)
Omission bias (Experiment 5)	Belief bias in syllogistic reasoning (Experiment 8; Stanovich & West, 1998c)
One-side bias, within subjects (Stanovich & West, in press)	Belief bias in modus ponens (Experiment 8)
Certainty effect (Experiment 6)	Informal argument evaluation (Experiment 8; Stanovich & West, 1997)
WTP/WTA difference (Experiment 6)	Four-card selection task (Experiment 8; Stanovich & West, 1998a)
Myside bias between and within subjects (Experiment 7; Stanovich & West, 2007, in press)	Expected value maximization in gambles (Benjamin & Shapiro, 2005; Frederick, 2005)
Newcomb's problem (Stanovich & West, 1999; Toplak & Stanovich, 2002)	Overconfidence effect (Bruine de Bruin et al., 2007; Stanovich & West, 1998c)

from Perkins' original coinage), refers to the rules, procedures, and strategies that can be retrieved by the analytic system and used to substitute for the heuristic response. However, if the mindware available to the analytic system for heuristic override has not been learned, then we have a case of a mindware gap.

In contrast, errors of application can only occur when the relevant mindware has been learned and is available for use in the override process. Errors of application occur when people fail to detect the situational cues indicating that the heuristically primed response needs to be overridden and an analytically derived response substituted. We give this requirement the label override detection (detecting the necessity for heuristic override). The above quote from Kahneman and Frederick (2002) suggests that cognitive ability differences only arise when the experimental task allows for variation in the presence of the relevant mindware and in the override detection process. It will be argued here that this analysis ignores a third potent source of nonnormative responding that might be an even more important source of individual differences.

A Framework for Individual Differences in Heuristics and Biases Tasks

Most of the tasks in the heuristics and biases literature were deliberately designed to pit a heuristically triggered response against a normative response generated by the analytic system. As Kahneman (2000) noted, "Tversky and I always thought of the heuristics and biases approach as a two-process theory" (p. 682). However, what this means is that even after the necessity for override has been detected and the relevant mindware is available, the conflict has to be resolved. Resolving the conflict in favor of the analytic response may require cognitive capacity, especially if cognitive decoupling must take place for a considerable period of time while the analytic response is computed. Cognitive decoupling is involved in inhibiting the heuristic response and also in simulating alternative responses (Stanovich, 2008a, 2008b). Recent work on inhibition and executive functioning has indicated that such cognitive decoupling is very capacity demanding and that it is strongly related to individual differences in fluid intelligence (Duncan, Emslie, Williams, Johnson, & Freer, 1996; Engle, 2002; Geary, 2005; Gray, Chabris, & Braver, 2003; Kane & Engle, 2002, 2003; Salthouse, Atkinson, & Berish, 2003; Unsworth & Engle, 2005, 2007).

It is argued here that it is this third factor present in some heuristics and biases tasks—the necessity for sustained cognitive decoupling—that is the major source of the variability in the association between cognitive ability and task performance that is displayed in Table 8. Building on the conjectures of Kahneman (2000) and Kahneman and Frederick (2002), our framework for conceptualizing individual differences on heuristics and biases tasks is displayed in Figure 1. The question addressed in the first stage of the framework is whether, for a given task, the mindware is available to carry out override (whether the procedures and declarative knowledge are available to substitute an analytic response for a heuristic one). If the relevant mindware is not available, then the person must, of necessity, respond heuristically. It is immaterial whether the person detects the necessity for override or has the capacity to sustain override if the normatively appropriate response is simply not available. If the relevant mindware (probabilistic thinking skills, falsifiability tendencies, disposition to search for alternative explanations, sensitivity to contradiction, etc.) is not present, then participants will end up at what has been termed in the figure Path 1 to a heuristic response.

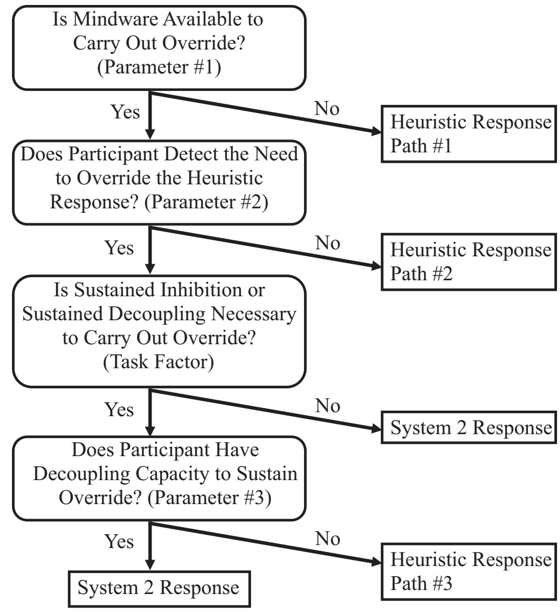

Figure 1. A framework for conceptualizing individual differences on heuristics and biases tasks.

If the relevant mindware is in fact available, then the next question that becomes operative is whether or not the person detects the need to override the heuristic response. Even if the relevant mindware is present, if the participant does not detect any reason to override the heuristic response, then it will be emitted (this is Path 2 to a heuristic response as labeled in the figure). Many heuristics and biases tasks lead people down this path. People do not detect the need to override the response that comes naturally (Kahneman, 2003) even though, in retrospect, they would endorse the norm that the heuristic response violated (Kahneman & Tversky, 1982a; Shafir, 1998; Shafir & Tversky, 1995; Thaler, 1987).

The next choice point in the figure concerns the task rather than the participant. If the relevant mindware is present and if the need for override has been noted, the question then becomes whether or not the task requires sustained inhibition (cognitive decoupling) in order to carry out the override of the heuristic response. If not (or if the capacity required is low—this of course may not be an all-or-nothing issue), then the analytic (System 2) response will be substituted for the heuristic response. In contrast, if the task requires sustained decoupling in order to carry out override, then we must ask whether the participant has the cognitive capacity that will be necessary. If so, then the analytic response will be given. If not, then the heuristic response will be given (Path 3 to the heuristic response in the figure)—despite the availability of the relevant mindware and the recognition of the need to use it.

In order for cognitive ability to associate with a bias, there must be differences correlated with cognitive ability at some of the choice points in the framework—that is, in some of the person parameters that branch toward or away from heuristic paths. As Kahneman (2000) noted, "A task will be too difficult if (1) System 1 favors an incorrect answer, and (2) System 2 is incapable of applying the correct rule, either because the rule is unknown [mindware gap] or because the cues that would evoke it are absent [no override detection]" (p. 682). Performance on such a task will be floored and will show no association with cognitive ability. Some of the tasks in Table 8 are no doubt of this type (between-subjects conjunction effects, for example). However, several of the tasks in Table 8 without associations with cognitive ability cannot be viewed as displaying floor effects.[2] For a cognitive ability difference to be observed, there must be differential cleaving by intelligence at some of the critical nodes in Figure 1—that is, there must be a correlation between intelligence and at least one of the person parameters. Of course, the partitioning of cognitive ability groups at each of the nodes will vary from task to task. We advance here a generic conjecture about the source of associations with cognitive ability. The conjecture is that the primary source of associations with cognitive ability in heuristics and biases tasks is the way that people are partitioned by person Parameter 3 ("does the person have the decoupling capacity to sustain override").

Cognitive Decoupling, Mindware Gaps, and Override Detection in Heuristics and Biases Tasks

There is evidence in the literature indicating that intelligence tests (especially tests of fluid intelligence, see Carroll, 1993; Horn & Cattell, 1967; Horn & Noll, 1997) directly tap the ability to sustaining the decoupling of representations from the world so that cognitive simulations can be run which test the outcomes of imaginary actions (Currie & Ravenscroft, 2002; Dienes & Perner, 1999; Evans & Over, 2004; Kahneman & Tversky, 1982b; Nichols & Stich, 2003). There is probably a substantial differential in the cleaving at Node 3 based on cognitive ability (Duncan et al., 1996; Kane & Engle, 2002; Salthouse et al., 2003). In contrast, we conjecture that, for many tasks in the heuristics and biases literature, the other two parameters show only modest differential partitioning based on cognitive ability.

Regarding Parameter 1, it is true that the rules, knowledge, and strategies available to the analytic system to use in heuristic system overrides are in part the product of past learning experiences. One might expect that people with more cognitive ability would profit more from learning experiences. However, the relevant mindware for our present discussion is not just generic procedural knowledge, nor is it the hodge-podge of declarative knowledge that is often used to assess crystallized intelligence on ability tests. Instead, it is a very special subset of knowledge related to how one views probability and chance; whether one has the tools to think scientifically and the propensity to do so; the tendency to think logically; and knowledge of some special rules of formal reasoning and good argumentation. At least among the university students typically tested in these studies, acquiring these sets of skills and knowledge bases might be, experientially, very haphazard.

Although it is true that more intelligent individuals learn more things than the less intelligent, many thinking dispositions relevant to rationality are acquired rather late in life and the explicit teaching of this mindware is very spotty and inconsistent. For example, the tendency to think of alternative explanations for a phenomenon leads to the ability to more accurately infer causal models of events. Such principles are taught very inconsistently (by either explicit or implicit means). Or take, for example, the conjunction rule of probability, the violation of which is illustrated in the Linda problem. Kahneman and Tversky (1982a) reported that tests of rule endorsement and argument endorsement conducted after participants had made the conjunction error revealed that statistically sophisticated psychology graduate students did endorse the rule they had violated (they possessed the relevant mindware but did not detect the necessity for override). However, a majority of statistically naïve undergraduate students failed to endorse the conjunction rule—they lacked the relevant mindware ("much to our surprise, naive subjects did not have a solid grasp of the conjunction rule," p. 127, Kahneman & Tversky, 1982a). The lack of uniform teaching and learning conditions for the acquisition of this mindware might attenuate any natural correlation with intelligence that there would be if it were taught under uniform conditions.

Override detection (Parameter 2), we would argue, is perhaps even more likely to display a low correlation with cognitive ability. First, it would seem to be more of a thinking disposition (related to constructs like need for cognition; e.g., see Cacioppo et al., 1996) than a cognitive capacity. Psychometricians have long distinguished typical performance situations from optimal (sometimes termed maximal) performance situations (Ackerman, 1994, 1996; Ackerman & Heggestad, 1997; 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. In contrast, optimal performance situations are those in which the task interpretation is determined externally (not left to the participant); the participant is instructed to maximize performance and is told how to do so. All tests of intelligence or cognitive aptitude are optimal performance assessments, whereas measures of rational thinking dispositions (need for cognition, actively openminded thinking, reflectivity/impulsivity) are often assessed under typical performance conditions.

Override detection, particularly in between-subjects designs, is exercised under typical rather than optimal conditions. It thus parses, in terms of the structure of cognitive abilities (Baron, 1985, 2000; Dole & Sinatra, 1998; Matthews et al., 2002; Sinatra & Pintrich, 2003; Sternberg, 1997, 2003) with thinking dispositions rather than cognitive capacity measures such as intelligence. For these theoretical reasons, we think that person Parameter 2 in the framework is less the source of associations with cognitive ability than is Parameter 3.

---

[2] Although the presence of floor and ceiling effects can be difficult to determine in between-subjects experiments, it is clear that at least some of our between-subjects tasks avoided them. For example, in the engineer/lawyer base-rate problem from Experiment 1, roughly half of the sample made estimates below 60% in the 30% engineer condition, even though the description tilts toward engineer. This seems to show some sensitivity to the base rate and to thus indicate that at least half the sample was not floored. It is likewise with the African countries anchoring problem of Experiment 1. Regardless of the anchor, more than half the sample gave estimates less than 55 and greater than 15. Given that the actual number is 53, such performance is indicative of at least some degree of resistance to anchoring in the task.

There are two other ways that the influence of Parameter 2, as a generator of individual differences, becomes attenuated—essentially by floor effects (as Kahneman, 2000, argued), but also by ceiling effects. Certain tasks in between-subjects designs (perhaps anchoring problems or the Linda problem) give so few cues to the possibility of heuristic/analytic conflict that this parameter is probably floored for most subjects. Conversely, the instructions in other tasks (belief bias assessed with syllogisms, for example), and some situations in real life ("the salesperson is trying to sell you—don't forget") are so explicit in calling attention to heuristic/analytic conflict that this parameter is probably near ceiling.

The case of belief bias in syllogistic reasoning is probably a good illustration of our argument that it is Parameter 3—the decoupling capacity parameter—that is the primary generator of associations with cognitive ability in rational thinking tasks[3] (see De Neys, 2006a, 2006b). The mindware available to reason logically on these simple categorical syllogisms (Parameter 1) is probably pretty uniformly present in the sample of university students studied here (and in most studies in the reasoning literature). The procedures needed to reason through the syllogisms used in these studies (for example, the invalid syllogism: all A are B, all C are B, therefore all C are A) are within the mindware of the vast majority of the students in research studies (the percentage correct in the neutral condition in Experiment 8 [84.1%] was almost as high as the percentage correct in the consistent condition [85.0%]). Additionally, as just mentioned, the instructions on this task probably ceiling out Parameter 2—override detection. Recall that the instructions to the task sensitize the participants to potential conflict (between argument validity and the truth of argument components). Thus, Parameters 1 and 2 probably leave little room for any individual difference variable to associate with performance.

In contrast, the task does require sustained cognitive decoupling (De Neys, 2006b). In the "rose" syllogism for example (All flowers have petals; roses have petals; therefore, roses are flowers—which is invalid), participants must suppress the tendency to endorse a valid response because of the "naturalness" (see Kahneman, 2003) of the conclusion—roses are flowers. This response must be held in abeyance while reasoning procedures work through the partially overlapping set logic indicating that the conclusion does not necessarily follow and that the syllogism is thus invalid. The reasoning process may take several seconds of perhaps somewhat aversive concentration (see Botvinick, Cohen, & Carter, 2004; Glenberg, 1997; Kahneman, 1973; Navon, 1989)—seconds during which the urge to foreclose the conflict by acceding to the natural tendency to affirm "roses are flowers" (by responding "valid") must be suppressed. Such response suppression while reasoning is closely related to the inhibitory and conflict resolution processes being studied by investigators examining the construct of executive functioning (Baddeley, Chincotta, & Adlam, 2001; Botvinick et al., 2004; Kane, 2003; Salthouse et al., 2003). Individual differences in such inhibitory processes have been found to be strongly associated with individual differences in fluid intelligence.

We conjectured that many of the other tasks that do show associations with cognitive ability (second column of Table 8) are tasks that involve some type of inhibition and/or sustained cognitive decoupling. For example, in within-subjects tests of outcome bias (Stanovich & West, 1998c) the appearance of the second item gives a pretty clear signal to the participant that there is an issue of consistency in their responses to the two different forms—that is, the within-subjects design probably puts Parameter 2 at ceiling, thus insuring that it is not the source of any associations with cognitive ability that are obtained. Detecting the need for consistency is not the issue. Instead, the difficulty comes from the necessity of inhibiting the tendency to downgrade the decision in the negative outcome condition, despite its having a better rationale than the positive outcome decision. Even in the between-subjects version of this task, one group of participants—those getting the negative outcome version—is alerted to the potential conflict between the seemingly good reasons to have the operation and the shockingly bad outcome. Participants must suppress the desire to sanction the decision, decouple their knowledge of the outcome, and simulate (see Evans & Over, 1996; Nichols & Stich, 2003) what they would have thought had they not known the outcome. Indeed, this condition creates a situation similar to those of various "curse of knowledge" paradigms (see Birch, 2005; Camerer, Loewenstein, & Weber, 1989; Gilovich, Medvec, & Sativsky, 1998; Hinds, 1999; Keysar & Barr, 2002; Royzman, Cassidy, & Baron, 2003). Note that the two cognitive ability groups show no difference in the positive outcome condition (see Table 1), which does not necessitate heuristic system override. The difference is entirely in the negative outcome condition in which sustained suppression is required.

The "curse of knowledge" logic of the negative item in the outcome bias task is similar to that in hindsight bias paradigms (e.g., Christiansen-Szalanski & Williams, 1991; Fischhoff, 1975; Pohl, 2004), which have also shown associations with cognitive ability (Stanovich & West, 1998c). In hindsight paradigms, the marking of the correct response sensitizes every respondent to the potential conflict involved—between what you know now versus what you would have known without the correct response being indicated. Thus again, Parameter 2 must be at ceiling. However, there is a need for sustained decoupling in the task, so whatever association between bias and cognitive ability exists on the task (a modest one; see Stanovich & West, 1998c) is likely generated by individual differences in Parameter 3.

Within-subjects framing paradigms probably have a similar logic. The appearance of the second problem surely signals that an issue of consistency is at stake (putting Parameter 2 at ceiling), and virtually all of the university students in these studies have acquired the value of consistency (Parameter 1 is also at ceiling). The modest cognitive ability associations that are generated by this task probably derive from lower cognitive ability participants who cannot suppress the attractiveness of an alternative response despite the threat to consistent responding that it represents—in short, from variation in Parameter 3. In contrast, between-subjects framing situations probably drive Parameter 2 to a very low value (few people recognize that there is a conflict to be resolved between a potentially different response to an alternative framing),

[3] Our conjecture here amounts to an endorsement of what Evans (in press) calls the quality hypothesis regarding cognitive ability—that individuals higher in cognitive ability are more likely to compute the correct response given that they have engaged System 2. The corresponding quantity hypothesis is that individuals higher in cognitive ability are more likely to engage System 2.

thus eliminating associations with individual differences (in the manner suggested by Kahneman, 2000).

The logic of the Linda problem is similar. Transparent, within-subjects versions are easier because they signal the conflict involved and the necessity for override. Such versions create at least modest associations with cognitive ability. In the between-subjects version, however, individual differences are eliminated entirely because this design obscures the heuristic/analytic conflict and puts Parameter 2 at floor.

As a final example, consider the difference between causal and noncausal base rates[4] illustrated in Table 8. Noncausal base-rate problems trigger conflict detection in so few participants that Parameter 2 is floored and hence cognitive ability differences are eliminated. In contrast, in a classic causal base-rate problem such as the Volvo versus Saab problem (see Footnote 4), where aggregate information is pitted against indicant information, the aggregate information has a causal relationship to the criterion behavior. Thus, the aggregate information in causal base-rate scenarios clearly signals that there are two pieces of information in conflict, Parameter 2 is near ceiling, and individual differences are determined largely by Parameter 3 (the sustained decoupling parameter), which is, we conjecture, linked to individual differences in cognitive ability. Thus, causal, but not noncausal, base-rate problems show cognitive ability differences.

Of course, with this discussion of what creates associations between biases and cognitive ability, we do not mean to draw attention away from the most salient outcome of this investigation—that a startlingly wide range of rational thinking tendencies appear to be independent of intelligence within the range existing in a university sample. These include many tasks that test some very basic strictures of rational thought. For example, the absence of framing and context effects are performance patterns that ensure that people's choices are utility maximizing. The failure to adhere to these strictures leads to descriptive models of human rationality that have profound public policy implications (Camerer, Issacharoff, Loewenstein, O'Donoghue, & Rabin, 2003; Mitchell, 2005; Prentice, 2005; Sunstein & Thaler, 2003). But adherence to these strictures of utility maximization were unrelated to cognitive ability in our sample. Within the range of intelligence that we studied, individuals of the highest cognitive capacity were no less likely to display base-rate neglect, the conjunction fallacy, myside bias, anchoring effects, the sunk-cost effect, proportion dominance, and a host of other cognitive biases.

General Conclusions

The framework in Figure 1 illustrates why rationality will not be uniformly related to intelligence. Instead, that relationship will depend upon the degree that rational responding requires sustained cognitive decoupling. When the heart of the task is recognizing the need for heuristic override but the override operation itself is easily accomplished, no sustained decoupling is necessary and rational thinking will depend more on the operations of the reflective mind than on those of the algorithmic mind (Stanovich, 2008a, 2008b). Thus, relationships with intelligence will be attenuated. Additionally, as Kahneman (2000) has argued, when detecting the necessity for override is very difficult (Parameter 2 is low), performance overall will be quite low and no relationships with cognitive ability will be evident.

Conversely, however, highly intelligent people will display fewer reasoning biases when you tell them what the bias is and what they need to do to avoid it. That is, when Parameters 1 and 2 are ceilinged and considerable cognitive capacity is needed to sustain decoupling while the correct response is computed, then highly intelligent people will do better in a rational thinking task. However, if there is no advance warning that biased processing must be avoided (as is the case in many between-subjects designs), then more intelligent individuals are not much more likely to perform any better on the task. Another way to phrase this is to say that, often, people of higher cognitive ability are no more likely to recognize the need for a normative principle than are individuals of lower cognitive ability. When the former believe that nothing normative is at stake, they behave remarkably like other people (equally likely for example to be "anchored" into responding that redwoods are almost 1,000 ft tall!—see Table 1). If told, however, that they are in a situation of normative conflict and if resolving the conflict requires holding a prepotent response in abeyance, then the individual of high cognitive ability will show less of many different cognitive biases.

An important caveat to the model presented in Figure 1 is that which rational thinking tasks yield a conflict between heuristic and analytic responses is not fixed, but instead is a function of the individual's history of mindware acquisition. Early in developmental history, the relevant mindware will not be present and the heuristic response will be inevitable—no conflict will even be detected. Someone with no training in thinking probabilistically—or, for that matter, logically in terms of subset and superset—may experience no conflict in the Linda problem. As experience with statistical and probabilistic thinking grows, a person will begin to experience more of a conflict because relevant mindware is available for use in the simulation of an alternative response by the analytic system. The final developmental stage in this sequence might well be that the mindware used in analytic simulation becomes so tightly compiled that it is triggered in the manner of a natural heuristic response. Some statistics instructors, for example, become unable to empathize with their students for whom the basic probability axioms are not transparent. The instructor can no longer remember when these axioms were not primary intuitions. This final stage of processing is perhaps captured by developmental models of heuristic versus analytic processing that trace a trajectory where fluent adult performance looks very heuristic (Brainerd & Reyna, 2001; Ericsson & Charness, 1994; Klein, 1998; Reyna, 2004; Reyna & Brainerd, 1995; Reyna, Lloyd, & Brainerd, 2003).

---

[4] Base rates that have a causal relationship to the criterion behavior (Ajzen, 1977; Bar-Hillel, 1980, 1990; Tversky & Kahneman, 1979) are often distinguished from noncausal base-rate problems—those involving base rates with no obvious causal relationship to the criterion behavior. A famous noncausal problem is the well-known cab problem (see Bar-Hillel, 1980; Lyon & Slovic, 1976; Tversky & Kahneman, 1982). In contrast, the classic Volvo versus Saab item (see p. 285 of Fong, Krantz, & Nisbett, 1986) would be an example of a causal base-rate 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, NJ: 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., Kyllonen, P., & Richards, R. (Eds.). (1999). *Learning and individual differences: Process, trait, and content determinants.* Washington, DC: American Psychological Association.

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology, 35,* 303–314.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axioms de l'scole americaine. [Rational man's behavior in face of risk: Critique of the American School's postulates and axioms]. *Econometrica, 21,* 503–546.

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin, 125,* 591–600.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35,* 124–140.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130,* 641–657.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44,* 211–233.

Bar-Hillel, M. (1990). Back to base rates. In R. M. Hogarth (Eds.), *Insights into decision making: A tribute to Hillel J. Einhorn* (pp. 200–216). Chicago: University of Chicago Press.

Baron, J. (1985). *Rationality and intelligence.* Cambridge, England: Cambridge University Press.

Baron, J. (1991). Beliefs about thinking. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 169–186). Hillsdale, NJ: Erlbaum.

Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making.* New York: Oxford University Press.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge, England: Cambridge University Press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54,* 569–579.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94,* 74–85.

Bartels, D. M. (2006). Proportion dominance: The generality and variability of favoring relative savings over absolute savings. *Organizational Behavior and Human Decision Processes, 100,* 76–95.

Bazerman, M., Baron, J., & Shonk, K. (2001). *"You can't enlarge the pie": Six barriers to effective government.* New York: Basic Books.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research, 30,* 961–981.

Benjamin, D., & Shapiro, J. (2005, February 25). *Does cognitive ability reduce psychological bias*? JEL manuscript: J24, D14, C91.

Birch, S. A. J. (2005). When knowledge is a curse: Children's and adult's reasoning about mental states. *Current Directions in Psychological Science, 14,* 25–29.

Botvinick, M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8,* 539–546.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. In H. W. Reese & R. Kail (Eds.), *Advances in child development and behavior* (Vol. 28, pp. 41–100). San Diego, CA: Academic 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.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review, 151,* 1211–1254.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy, 97,* 1232–1254.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.

Cassady, J. C. (2001). Self-reported GPA and SAT: A methodological note. *Practical Assessment, Research & Evaluation, 7*(12). Retrieved October 10, 2006, from http://www.pareonline.net/getvn.asp?v=7&n=12

Christiansen-Szalanski, J. J., & Williams, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes, 48,* 147–168.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science.* New York: Oxford University Press.

College Board. (2006). *College Board SAT: 2006 College-bound seniors: Total group profile report.* Retrieved March 23, 2007, from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2006/national-report.pdf

Cronbach, L. J. (1949). *Essentials of psychological testing.* New York: Harper.

Currie, G., & Ravenscroft, I. (2002*). Recreative minds: Imagination in philosophy and psychology.* Oxford, England: Clarendon Press.

Dawes, R. M. (1998). Behavioral decision making and judgment. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 497–548). Boston: McGraw-Hill.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford, England: Oxford University Press.

Deary, I. J. (2001). *Intelligence: A very short introduction.* Oxford, England: Oxford University Press.

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. (2006a). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology, 59,* 1070–1100.

De Neys, W. (2006b). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17,* 428–433.

DeShon, R. P., Smith, M. R., Chan, D., & Schmitt, N. (1998). Can racial differences in cognitive test performance be reduced by presenting problems in a social context? *Journal of Applied Psychology, 83,* 438–451.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22,* 735–808.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist, 33,* 109–128.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30,* 257–303.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51,* 380–417.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23.

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin, 30,* 447–460.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725–747.

Evans, J. St. B. T. (2002). The influence of prior belief on scientific thinking. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 193–210). Cambridge, England: Cambridge University Press.

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. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review, 13,* 378–395.

Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment.* New York: Psychology Press.

Evans, J. St. B. T. (in press). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning.*

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., Newstead, S. E., Allen, J., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology, 6,* 263–285.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction.* Hove, England: Erlbaum.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning.* Hove, England: Psychology Press.

Evans, J. St. B. T., & Over, D. E. (2004). *If.* Oxford, England: Oxford University Press.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13,* 1–17.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance, 23,* 339–359.

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.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science, 15,* 373–378.

Gal, I., & Baron, J. (1996). Understanding repeated simple choices. *Thinking and Reasoning, 2,* 81–98.

Geary, D. C. (2005). *The origin of the mind: Evolution of brain, cognition, and general intelligence.* Washington, DC: American Psychological Association.

George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86,* 93–111.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge University Press.

Gilovich, T., Medvec, V. H., & Sativsky, K. (1998). The illusion of transparency: Biased assessment of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75,* 332–346.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1–55.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6,* 316–322.

Greenhoot, A. F., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology, 18,* 203–221.

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 and Reasoning, 10,* 175–195.

Higgins, D. M., Peterson, J. B., Pihl, R. O., & Lee, A. G. M. (2007). Prefrontal cognitive ability, intelligence, big five personality, and the prediction of advanced academic and workplace performance. *Journal of Personality and Social Psychology, 93,* 298–319.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on predictions of novice performance. *Journal of Experimental Psychology: Applied, 5,* 205–221.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 1–23.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. Flanagan, J. Genshaft, & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin, 125,* 576–590.

Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology, 86,* 680–695.

Jeffrey, R. C. (1983). *The logic of decision* (2nd ed.). Chicago: University of Chicago Press.

Johnson-Laird, P. N. (2006). *How we reason.* Oxford, England: Oxford University Press.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

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. A. (2003). 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., & Frederick, S. (2005). A model of intuitive judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kahneman, D., & Tversky, A. (1982a). On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 493–508). Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982b). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341–350.

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.* New York: Cambridge University Press.

Kane, M. J. (2003). The intelligent brain in conflict. *Trends in Cognitive Sciences, 7,* 375–377.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review, 9,* 637–671.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132,* 47–70.

Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users do not do what they "should." In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 150–166). New York: 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.

Klaczynski, P. A., & Lavallee, K. L. (2005). Domain-specific identity, epistemic regulation, and intellectual ability as predictors of belief-based reasoning: A dual-process perspective. *Journal of Experimental Child Psychology, 92,* 1–24.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107,* 852–884.

Klein, G. (1998). *Sources of power: How people make decisions.* Cambridge, MA: MIT Press.

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.

Kuhn, D. (2001). How do people know? *Psychological Science, 12,* 1–8.

Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75,* 63–82.

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making, 16,* 77–92.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92,* 805–824.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points." *Annual Review of Psychology, 51,* 405–444.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General Intelligence,' Objectively Determined and Measured." *Journal of Personality and Social Psychology, 86,* 96–111.

Lubinski, D., & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence, 24,* 159–201.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions.* New York: Wiley.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica, 40,* 287–298.

Maher, P. (1993). *Betting on theories.* Cambridge, England: Cambridge University Press.

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

Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science & myth.* Cambridge, MA: MIT Press.

Mitchell, G. (2005). Libertarian paternalism is an oxymoron. *Northwestern University Law Review, 99,* 1245–1278.

Navon, D. (1989). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system. *European Journal of Cognitive Psychology, 1,* 191–238.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology, 57A,* 33–60.

Nichols, S., & Stich, S. (2003). *Mindreading.* Oxford, England: Oxford University Press.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93,* 116–130.

Over, D. E. (2004). Rationality and the normative/descriptive distinction. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 3–18). Malden, MA: Blackwell Publishing.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76,* 972–987.

Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual differences approach. *Journal of Behavioral Decision Making, 18,* 1–27.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77,* 562–571.

Perkins, D. N. (1995). *Outsmarting IQ: The emerging science of learnable intelligence.* New York: Free Press.

Perkins, D. N. (2002). The engine of folly. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 64–85). New Haven, CT: Yale University Press.

Perkins, D. N., & Ritchhart, R. (2004). When is good thinking? In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition* (pp. 351–384). Mahwah, NJ: Erlbaum.

Platt, R. D., & Griggs, R. A. (1993). Facilitation in the abstract selection task: The effects of attentional and instructional factors. *Quarterly Journal of Experimental Psychology, 46A,* 591–613.

Platt, R. D., & Griggs, R. A. (1995). Facilitation and matching bias in the abstract selection task. *Thinking and Reasoning, 1,* 55–70.

Pohl, R. (2004). Hindsight bias. In R. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment and memory* (pp. 363–378). Hove, England: Psychology Press.

Pollard, P., & Evans, J. St. B. T. (1987). Content and context effects in reasoning. *American Journal of Psychology, 100,* 41–60.

Prentice, R. A. (2005). Chicago Man, K-T Man, and the future of behavioral law and economics. *Vanderbilt Law Review, 56,* 1663–1777.

Reyna, V. F. (2004). How people make decisions that involve risk. *Current Directions in Psychological Science, 13,* 60–66.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75.

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.

Royzman, E. B., Cassidy, K. W., & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology, 7,* 38–65.

Sá, W., Kelley, C., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: Individual differences in the coordination of theory and evidence. *Personality and Individual Differences, 38,* 1149–1161.

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.

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, England: Oxford University Press.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Schick, F. (1987). Rationality: A third dimension. *Economics and Philosophy, 3,* 49–66.

Shafir, E. (1998). Philosophical intuitions and cognitive mechanisms. In M. R. DePaul & W. Ramsey (Eds.), *Rethinking intuition* (pp. 59–83). Lanham, MD: Rowman & Littlefield.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53,* 491–517.

Shafir, E., & Tversky, A. (1995). Decision making. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol. 3, pp. 77–100). Cambridge, MA: The MIT Press.

Sinatra, G. M., & Pintrich, P. R. (Eds.). (2003). *Intentional conceptual change.* Mahwah, NJ: Erlbaum.

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

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York: Cambridge University Press.

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science, 15,* 322–325.

Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science, 19,* 368–373.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57,* 31–95.

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. (2008a). 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.* Oxford, England: Oxford University Press.

Stanovich, K. E. (2008b). *Rationality and the reflective mind: Toward a tri-process model of cognition.* 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 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.

Stanovich, K. E., & West, R. F. (in press). On the failure of intelligence to predict myside bias and one-sided bias. *Thinking & Reasoning.*

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning.* Hillsdale, NJ: Lawrence Erlbaum.

Sternberg, R. J. (1997). *Thinking styles.* Cambridge, England: Cambridge University Press.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence.* Cambridge, England: Cambridge University Press.

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized.* Cambridge, England: Cambridge University Press.

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review, 70,* 1159–1202.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1,* 39–60.

Thaler, R. H. (1987). The psychology of choice and the assumptions of economics. In A. E. Roth (Ed.), *Laboratory experimentation in economics: Six points of view* (pp. 99–130). Cambridge, England: University of Cambridge Press.

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.

Toplak, M., & 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.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Tversky, A., & Kahneman, D. (1979). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology.* Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). Cambridge, England: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* 251–278.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics, 106,* 1039–1061.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33,* 67–81.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132.

U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. (2003). *Trends in the well-being* *of America's children and youth 2003*. Retrieved February 25, 2007, from http://aspe.hhs.gov/hsp/03trends/index.htm

Valentine, E. R. (1975). Performance on two reasoning tasks in relation to intelligence, divergence and interference proneness: Content and context effects in reasoning. *British Journal of Educational Psychology, 45,* 198–205.

von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

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

West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science, 2,* 325–330.

West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition, 31,* 243–251.

Received May 23, 2007
Revision received November 1, 2007
Accepted November 12, 2007 ■