

Cognitive Reflection and Decision Making

Author(s): Shane Frederick

Source: *The Journal of Economic Perspectives*, Vol. 19, No. 4 (Autumn, 2005), pp. 25-42

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/4134953>

Accessed: 18-01-2018 17:58 UTC

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

*American Economic Association* is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Economic Perspectives*

## Cognitive Reflection and Decision Making

Shane Frederick

**P**eople with higher cognitive ability (or “IQ”) differ from those with lower cognitive ability in a variety of important and unimportant ways. On average, they live longer, earn more, have larger working memories, faster reaction times and are *more* susceptible to visual illusions (Jensen, 1998). Despite the diversity of phenomena related to IQ, few have attempted to understand—or even describe—its influences on judgment and decision making. Studies on time preference, risk preference, probability weighting, ambiguity aversion, endowment effects, anchoring and other widely researched topics rarely make any reference to the possible effects of cognitive abilities (or cognitive *traits*).

Decision researchers may neglect cognitive ability because they are more interested in the *average* effect of some experimental manipulation. On this view, individual differences (in intelligence or anything else) are regarded as a nuisance—as just another source of “unexplained” variance. Second, most studies are conducted on college undergraduates, who are widely perceived as fairly homogenous. Third, characterizing performance differences on cognitive tasks requires terms (“IQ” and “aptitudes” and such) that many object to because of their association with discriminatory policies. In short, researchers may be reluctant to study something they do not find interesting, that is not perceived to vary much within the subject pool conveniently obtained, and that will just get them into trouble anyway.

But as Lubinski and Humphreys (1997) note, a neglected aspect does not cease to operate because it is neglected, and there is no good reason for ignoring the *possibility* that general intelligence or various more specific cognitive abilities are important causal determinants of decision making. To provoke interest in this

■ *Shane Frederick is Assistant Professor of Management Science, Sloan School of Management, Massachusetts Institute of Technology, Cambridge Massachusetts. His e-mail address is <shanefre@mit.edu>.*

neglected topic, this paper introduces a three-item “Cognitive Reflection Test” (CRT) as a simple measure of one type of cognitive ability. I will show that CRT scores are predictive of the types of choices that feature prominently in tests of decision-making theories, like expected utility theory and prospect theory. Indeed, the relation is sometimes so strong that the preferences themselves effectively function as expressions of cognitive ability—an empirical fact begging for a theoretical explanation.

After introducing the CRT, I examine its relations with two important decision-making characteristics: time preference and risk preference. The CRT is then compared with other measures of cognitive ability or cognitive “style,” including the Wonderlic Personnel Test (WPT), the Need For Cognition scale (NFC) and self-reported SAT and ACT scores. The CRT exhibits considerable difference between men and women, and I discuss how this relates to sex differences in time and risk preferences. The final section discusses the interpretation of correlations between cognitive abilities and decision-making characteristics.

## The Cognitive Reflection Test (CRT)

Many researchers have emphasized the distinction between two types of cognitive processes: those executed quickly with little conscious deliberation and those that are slower and more reflective (Epstein, 1994; Sloman, 1996; Chaiken and Trope, 1999; Kahneman and Frederick, 2002). Stanovich and West (2000) called these “System 1” and “System 2” processes, respectively. System 1 processes occur spontaneously and do not require or consume much attention. Recognizing that the face of the person entering the classroom belongs to your math teacher involves System 1 processes—it occurs instantly and effortlessly and is unaffected by intellect, alertness, motivation or the difficulty of the math problem being attempted at the time. Conversely, finding  $\sqrt{19163}$  to two decimal places without a calculator involves System 2 processes—mental operations requiring effort, motivation, concentration, and the execution of learned rules.<sup>1</sup>

The problem  $\sqrt{19163}$  allows no role for System 1. No number spontaneously springs to mind as a possible answer. Someone with knowledge of an algorithm and the motivation to execute it can arrive at the exact answer (138.43), but the problem offers no intuitive solution.

By contrast, consider this problem:

A bat and a ball cost \$1.10. The bat costs \$1.00 more than the ball.  
How much does the ball cost? \_\_ cents

Here, an intuitive answer *does* spring quickly to mind: “10 cents.” But this “impulsive” answer is wrong. Anyone who reflects upon it for even a moment would

<sup>1</sup> For a discussion of the distinction between System 1 and System 2 in the context of choice heuristics, see Frederick (2002).

Figure 1

**The Cognitive Reflection Test (CRT)**

- (1) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? \_\_\_\_\_ cents
- (2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? \_\_\_\_\_ minutes
- (3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? \_\_\_\_\_ days

recognize that the difference between \$1.00 and 10 cents is only 90 cents, not \$1.00 as the problem stipulates. In this case, catching that error is tantamount to solving the problem, since nearly everyone who does not respond “10 cents” does, in fact, give the correct response: “5 cents.”

In a study conducted at Princeton, which measured time preferences using both real and hypothetical rewards, those answering “10 cents” were found to be significantly less patient than those answering “5 cents.”<sup>2</sup> Motivated by this result, two other problems found to yield impulsive erroneous responses were included with the “bat and ball” problem to form a simple, three-item “Cognitive Reflection Test” (CRT), shown in Figure 1. The three items on the CRT are “easy” in the sense that their solution is easily understood when explained, yet reaching the correct answer often requires the suppression of an erroneous answer that springs “impulsively” to mind.

The proposition that the three CRT problems generate an incorrect “intuitive” answer is supported by several facts. First, among all the possible wrong answers people could give, the posited intuitive answers (10, 100 and 24) dominate. Second, even among those responding correctly, the wrong answer was often considered first, as is apparent from introspection, verbal reports and scribbles in the margin (for example, 10 cents was often crossed out next to 5 cents, but never the other way around). Third, when asked to judge problem difficulty (by estimating the proportion of *other* respondents who would correctly solve them), respondents who missed the problems thought they were easier than the respondents who solved them. For example, those who answered 10 cents to the “bat and ball” problem estimated that 92 percent of people would correctly solve it, whereas those who answered “5 cents” estimated that “only” 62 percent would. (Both were considerable overestimates.) Presumably, the “5 cents” people had mentally crossed out 10 cents and knew that not everyone would do this, whereas the “10 cents” people

<sup>2</sup> The “bat and ball” problem was subsequently used by Nagin and Pogarsky (2003) in a laboratory experiment on cheating. When respondents could obtain a \$20 reward for correctly answering six trivia questions, those answering 10 cents were significantly more likely to defy the experimenter’s request to complete the task without looking at the answers.

thought the problem was too easy to miss. Fourth, respondents do much better on analogous problems that invite more computation. For example, respondents miss the “bat and ball” problem far more often than they miss the “banana and bagel” problem: “A banana and a bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost?”

The CRT was administered to 3,428 respondents in 35 separate studies over a 26-month period beginning in January 2003. Most respondents were undergraduates at various universities in the midwest and northeast who were paid \$8 to complete a 45-minute questionnaire that included the CRT and measures of various decision-making characteristics, like time and risk preferences.<sup>3</sup> On the page on which the CRT appeared, respondents were told only: “Below are several problems that vary in difficulty. Try to answer as many as you can.”

Table 1 shows the mean scores at each location and the percentage answering 0, 1, 2 or 3 items correctly. Most of the analyses that follow compare the “low” group (those who scored 0 out of 3) with the “high” group (those who scored 3 out of 3). The two “intermediate” groups (those who scored a 1 or 2) typically fell between the two extreme groups on whatever dependent measure was analyzed. Thus, focusing attention on the two “extreme” groups simplifies the exposition and analysis without affecting the conclusions.

Since more of the respondents were college students from selective schools, the two “extreme” groups that formed the basis for most statistical comparisons were far more similar in cognitive abilities than two extreme groups formed from the general population. Thus, the group differences reported here likely *understate* the differences that would have been observed if a more representative sample had been used.

## Cognitive Reflection and Time Preferences

The notion that more intelligent people are more patient—that they devalue or “discount” future rewards less—has prevailed for some time. For example, in his *New Principles of Political Economy* (1834, pp. 57), Rae writes: “The strength of the intellectual powers, giving rise to reasoning and reflective habits. . . brings before us the future. . . in its legitimate force, and urge the propriety of providing for it.”

The widely presumed relation between cognitive ability and patience has been tested in several studies, although rather unsystematically. Melikian (1959) asked children from five to twelve years of age to draw a picture of a man, which they could exchange for either 10 fils (about 3 cents) or for a “promissory note” redeemable for 20 fils two days later. Those who opted for the promissory note scored slightly higher on an intelligence test based on an assessment of those

<sup>3</sup> There were three exceptions to this: 1) the participants from Carnegie Mellon University completed the survey as part of class; 2) the 4<sup>th</sup> of July participants received “only” a frozen ice cream bar; and 3) the participants from the web study were unpaid, although they were entered into a lottery for iPods and other prizes.

**Table 1**  
**CRT Scores, by Location**

Locations at which data were collected	Mean CRT score	Percentage scoring 0, 1, 2 or 3				N =
		"Low" 0	1	2	"High" 3	
Massachusetts Institute of Technology	2.18	7%	16%	30%	48%	61
Princeton University	1.63	18%	27%	28%	26%	121
Boston fireworks display <sup>a</sup>	1.53	24%	24%	26%	26%	195
Carnegie Mellon University	1.51	25%	25%	25%	25%	746
Harvard University <sup>b</sup>	1.43	20%	37%	24%	20%	51
University of Michigan: Ann Arbor	1.18	31%	33%	23%	14%	1267
Web-based studies <sup>c</sup>	1.10	39%	25%	22%	13%	525
Bowling Green University	0.87	50%	25%	13%	12%	52
University of Michigan: Dearborn	0.83	51%	22%	21%	6%	154
Michigan State University	0.79	49%	29%	16%	6%	118
University of Toledo	0.57	64%	21%	10%	5%	138
Overall	1.24	33%	28%	23%	17%	3428

Notes: <sup>a</sup> Respondents in this study were people picnicking along the banks of the Charles River prior to the July 4<sup>th</sup> fireworks display. Their ages ranged from 15 to 63, with a mean of 24. Many of the younger participants were presumably students at a college in the Boston or Cambridge area. Most completed the survey in small groups of friends or family. Although they were requested not to discuss it until everyone in their group had completed it, some may have. (This, presumably, would elevate the CRT scores relative to most of the other studies in which participation was more closely supervised.)

<sup>b</sup> The participants in this study were all members of a student choir group, which was predominately female. Unlike the other locations in which the numbers of men and women were comparable, 42 of 51 participants in this study were women.

<sup>c</sup> These were participants in two online studies, consisting of both college students and others whose e-mail addresses were obtained from online retailers.

drawings.<sup>4</sup> Funder and Block (1989) paid 14 year-olds to participate in six experimental sessions. For each of the first five sessions, they could choose between receiving \$4 or foregoing ("investing") their \$4 payment for \$4.80 in the sixth and final session. The teenagers with higher IQs chose to invest more of their money. In a follow-up to an extensive series of experiments investigating the ability of preschool children to delay gratification (Mischel, 1974), Shoda, Mischel and Peake (1990) found that the children who had waited longer before succumbing to the impulse to take an immediately available inferior reward scored higher on their SATs taken over a decade later. Similarly, Parker and Fischhoff (2005) found that scores on a vocabulary test taken around age eleven predicted the individual's tendency, at around age 18, to prefer a larger later reward over a smaller sooner one (for example, \$120 in four weeks to \$100 tomorrow). Using small real rewards, Benjamin and Shapiro (2005) found that respondents with higher SAT math scores

<sup>4</sup> Given the relatively wide range of ages in this study, it remains unclear whether this relation is attributable to intelligence, *per se*, or to age, which might correlate with the development of artistic skill or patience or trust or some other specific trait that can be distinguished from cognitive ability.

(or their Chilean equivalent) were more likely to choose a larger later reward over a smaller sooner one (for example, to prefer a postdated check for \$5.05 over a \$5.00 check that can be immediately cashed). However, Monterosso et al. (2001) found no relation between the IQ of cocaine addicts and their imputed discount rates, and Kirby, Winston and Santiesteban (2005) found no reliable relation between students' SAT scores and the amount they would bid for a delayed monetary reward (although they did find that college grade point averages correlated positively with those bids).

Collectively, these studies support the view that cognitive ability and time preference are somehow connected, though they have not generally focused on the types of intertemporal decisions over which cognitive ability exerts influence, nor explained why it does so.<sup>5</sup> Toward this end, I examined the relation between CRT scores and various items intended to measure different aspects of "time preference." As shown in Table 2, these included several hypothetical choices between an immediate reward and a larger delayed reward (items a through e), an immediate reward and a sequence of delayed rewards (items f through h), a shorter more immediate massage and longer more delayed massage (item i) and a smaller immediate loss or a larger delayed loss (items j and k).<sup>6</sup> Item l asked respondents to state their maximum willingness to pay to have a book shipped overnight rather than waiting two weeks. Item m involved real money. Through a series of choices, respondents specified the smallest amount of money in four days that they would prefer to \$170 in two months, and one of them was selected to actually receive one of their choices. Items n through q asked respondents to report their impulsivity, procrastination, preoccupation with their future and concerns about inflation on an 11-point scale ranging from -5 (much less than the average person taking this survey today) to +5 (much more than the average person taking this survey today).<sup>7</sup>

Table 2 shows the responses of the low and high CRT groups for each of the 17 items. The reported value is either the percentage choosing the patient option or the mean response. The subscripts are the total number of respondents in the low and high CRT groups who answered that item. The rightmost column reports the level of statistical significance of group differences—the p-values from a chi-square test (for dichotomous responses) or a t-test (for continuous responses).

Those who scored higher on the CRT were generally more "patient"; their decisions implied lower discount rates. For short-term choices between monetary rewards, the high CRT group was much more inclined to choose the later larger

<sup>5</sup> Shoda, Mischel and Peake (1990) examined preschoolers' willingness to wait (for additional marshmallows and pretzels and such) under four experimental conditions. They found that patience predicted SAT scores in only one of their four conditions—when the attractive but inferior reward was visually exposed and no distraction technique (such as "think fun") was suggested. In the other three conditions, patient behavior was actually *negatively* correlated with subsequent SAT scores.

<sup>6</sup> I assumed that delaying the extraction of a tooth involved a *larger* delayed loss, because during the intervening two weeks, one will suffer additional toothache pain, or additional disutility from dreading the forthcoming extraction pain, and that the only reason for *not* doing it immediately was that future pain was discounted relative to immediate pain.

<sup>7</sup> Among the items in Table 2, men were more patient for items c, k and l, and they worried more about inflation. There were no significant differences between men and women for any other item.

*Table 2*  
**Intertemporal Behavior for Low and High CRT Groups**  
*(percentage choosing patient option or mean response)*

<i>Item</i>	<i>Intertemporal Choice or Judgment</i>	<i>CRT group</i>		<i>Stat. Signif.</i>
		<i>Low</i>	<i>High</i>	
a	\$3400 this month or \$3800 next month	35% <sub>611</sub>	60% <sub>196</sub>	$p < 0.0001$
b	\$100 now or \$140 next year	22% <sub>409</sub>	37% <sub>297</sub>	$p < 0.0001$
c	\$100 now or \$1100 in 10 years	47% <sub>283</sub>	57% <sub>208</sub>	$p < 0.05$
d	\$9 now or \$100 in 10 years	40% <sub>364</sub>	46% <sub>277</sub>	$p < 0.10$
e	\$40 immediately or \$1000 in 10 years	50% <sub>135</sub>	59% <sub>83</sub>	n.s.
f	\$100 now or \$20 every year for 7 years	28% <sub>60</sub>	43% <sub>28</sub>	n.s.
g	\$400 now or \$100 every year for 10 years	64% <sub>44</sub>	72% <sub>43</sub>	n.s.
h	\$1000 now or \$100 every year for 25 years	52% <sub>295</sub>	49% <sub>99</sub>	n.s.
i	30 min. massage in 2 weeks or 45 min. massage in Nov.	28% <sub>272</sub>	27% <sub>126</sub>	n.s.
j	Lose \$1000 this year or lose \$2000 next year	78% <sub>166</sub>	73% <sub>86</sub>	n.s.
k	Tooth pulled today or tooth pulled in 2 weeks	59% <sub>430</sub>	65% <sub>242</sub>	n.s.
l	Willingness to pay for overnight shipping of chosen book	\$4.54 <sub>150</sub>	\$2.18 <sub>163</sub>	$p < 0.0001$
m	Smallest amount in 4 days preferred to \$170 in 2 months	\$116 <sub>72</sub>	\$133 <sub>82</sub>	$p < 0.01$
n	How impulsive are you?	+1.01 <sub>110</sub>	-0.21 <sub>47</sub>	$p < 0.001$
o	How much do you tend to procrastinate?	+1.05 <sub>110</sub>	+1.06 <sub>47</sub>	n.s.
p	How much do you think about your future?	+2.49 <sub>110</sub>	+1.64 <sub>47</sub>	$p < 0.01$
q	How much do you worry about inflation?	-1.16 <sub>110</sub>	+0.11 <sub>47</sub>	$p < 0.01$

reward (see items a and b). However, for choices involving longer horizons (items c through h), temporal preferences were weakly related or unrelated to CRT scores.

A tentative explanation for these results is as follows: a thoughtful respondent can find good reasons for discounting future monetary outcomes at rates exceeding the prevailing interest rate—the promiser could default, one may be predictably wealthier in the future (with correspondingly diminished marginal utility for further wealth gains), interest rates could increase (which increases the opportunity cost of foregoing the immediate reward), and inflation could reduce the future rewards' real value (if the stated amounts are interpreted as being denominated in nominal units).<sup>8</sup> Collectively, these reasons could, for example, justify choosing \$9 now over \$100 in 10 years (item d), even though the implied discount rate of such a choice (27 percent), exceeds market interest rates. However, such reasons are not sufficiently compelling to justify choosing \$3400 this month over \$3800 next month (which implies an annual discount rate of 280 percent). Hence, one observes considerable differences between CRT groups for choices like those in items a and b, where more careful deliberation or “cognitive reflection” should argue strongly in favor of the later larger reward, but negligible differences for many of the other items, for which additional reflection would not make such a strong case for the larger later reward (although one might argue that additional reflection should

<sup>8</sup> Frederick, Loewenstein and O'Donoghue (2002) offer a detailed and extended discussion of the conceptual dissection of imputed discount rates and discuss many reasons why choices between monetary rewards are problematic for measuring pure time preference.



reveal the wisdom of choosing the delayed 45-minute massage, since one will likely still be alive, still be stressed and sore, still like massages, and still derive greater benefits from longer ones).

It appears that greater cognitive reflection fosters the recognition or appreciation of considerations favoring the later larger reward (like the degree to which the implied interest rate exceeds the rate offered by the market). However, it remains unclear whether cognitive reflection also influences other determinants of intertemporal choices (like *pure* time preference). CRT scores were unrelated to preferences for the massage and tooth-pull items, which were intended as measures of pure time preference. On the other hand, those in the low CRT group (the “cognitively impulsive”) were willing to pay significantly more for the overnight shipping of a chosen book (item 1), which *does* seem like an expression of an aspect of *pure* time preference (the psychological “pain” of waiting for something desired).

Thus, despite the wide variety of items included to help address this issue, further resolution of the types of psychological characteristics associated with cognitive reflection (and other cognitive abilities) is still required. Toward this goal, respondents in some of the later studies were also asked to report several personality characteristics that seemed relevant to intertemporal choices (items n through q). The self-perceived tendency to procrastinate was unrelated to CRT scores (both groups thought that they procrastinate more than their peers). However, the high CRT group perceived themselves to be significantly *less* impulsive, *more* concerned about inflation and (curiously) *less* preoccupied with their future. The inflation result supports the idea that the high-scoring groups are more likely to consider such background factors in their choices between temporally separated monetary rewards. Its interpretation, however, is ambiguous, since it implies a consideration of future conditions, but would be a justification for choosing the proximate reward.

## Cognitive Reflection and Risk Preferences

In the domain of risk preferences, there is no widely shared presumption about the influences of cognitive ability and almost no research on the topic. Donkers, Melenberg and van Soest (2001) found that more educated respondents were more tolerant of risk in hypothetical gambles: for example, they were more likely to prefer an 80 percent chance of 45 florins (about \$23) over a sure 30 florins (about \$15). Benjamin and Shapiro (2005) found that students with higher scores on the math section of the SAT (or its Chilean equivalent) were more likely to choose according to expected value for real decisions involving small stakes (for example, they were more likely to prefer a 50 percent chance to win \$1.05 over a sure 50 cents).

To assess the relation between CRT and risk preferences, I included several measures of risk preferences in my questionnaires, including choices between a certain gain (or loss) and some probability of a larger gain (or loss). For some

items, expected value was maximized by choosing the gamble, and for some it was maximized by choosing the certain outcome.

The results are shown in Table 3a. In the domain of gains, the high CRT group was more willing to gamble—particularly when the gamble had higher expected value (top panel), but, notably, even when it did not (middle panel). If all five items from the middle panel of Table 3a are aggregated, the high CRT group gambled significantly more often than the low CRT group (31 percent versus 19 percent;  $\chi^2 = 8.82$ ;  $p < 0.01$ ). This suggests that the correlation between cognitive ability and risk taking in gains is not due solely to a greater disposition to compute expected value or to adopt that as the choice criterion.<sup>9</sup> For items involving losses (lower panel), the high CRT group was *less* risk seeking; they were more willing to accept a sure loss to avoid playing a gamble with lower (more negative) expected value.

Two pairs of items (d versus o and h versus r) were reflections of one another in the domain of gains and losses. Prospect theory predicts that people will be more willing to take risks to avoid losses than to achieve gains; that respondents will switch from risk aversion to risk seeking when the valence of a gamble (or “prospect”) changes from positive to negative (Kahneman and Tversky, 1979). Though this is spectacularly true for the low CRT group, who are much more willing to gamble in the domain of losses than in the domain of gains, there is no such reflection effect among the high CRT group, as shown in Table 3b. This result starkly shows the importance of considering cognitive ability when evaluating the descriptive validity of a theory of decision making.<sup>10</sup>

## Is the CRT Just Another IQ test?

Of the 3,428 respondents who completed the three-item CRT, many also completed one or more additional cognitive measures: 921 completed the Wonderlic Personnel Test (WPT)—a 12-minute, 50-item test used by the National

<sup>9</sup> As expected, the gamble was not popular among *either* group for *any* of the “anti-expected-value” gambles, since risk aversion and expected value both militate against it. However, any factors favoring the gamble over the sure thing (for example, valuing the excitement of gambling or dismissing the sure amount as negligibly small) would be more likely to tip preferences in favor of the gamble among those less averse to it (the high CRT group, as judged from items a through h). The gambles in items i through m were designed, in part, to have some chance of being chosen (the sure amounts were small, and the expected values of the gambles were typically close to the sure amount). Including choices in which the gambles lacked these properties (for example, offering a choice between \$4,000 for sure and a 50 percent chance of \$5000) would be pointless, because nearly everyone would reject the gamble, leaving no response variance to analyze. Item i comes close to illustrating this point.

<sup>10</sup> Although the descriptive accuracy of expected utility theory markedly *improves* for respondents with higher scores, it cannot explain why a 75 percent chance of \$200 is frequently rejected in favor of a sure \$100, across all levels of cognitive ability, since this is a small fraction of one’s wealth, and even a concave utility function is approximately linear over small changes (Rabin, 2000).

*Table 3a*  
**Risk Seeking Behavior among Low and High CRT Groups**

<i>Item</i>	<i>Percentage choosing riskier option</i>	<i>CRT group</i>		<i>Stat. Signif.</i>
		<i>Certain gains vs. Higher expected value gambles</i>	<i>Low</i>	
a	\$1,000 for sure or a 90% chance of \$5,000	52% <sub>280</sub>	74% <sub>225</sub>	$p < 0.0001$
b	\$100 for sure or a 90% chance of \$500	56% <sub>95</sub>	78% <sub>92</sub>	$p < 0.01$
c	\$1,000 for sure or a 75% chance of \$4,000	37% <sub>264</sub>	57% <sub>102</sub>	$p < 0.001$
d	\$100 for sure or a 75% chance of \$200	19% <sub>843</sub>	38% <sub>475</sub>	$p < 0.0001$
e	\$100 for sure or a 75% chance of \$150	10% <sub>217</sub>	34% <sub>94</sub>	$p < 0.0001$
f	\$100 for sure or a 50% chance of \$300	47% <sub>68</sub>	75% <sub>20</sub>	$p < 0.05$
g	\$500 for sure or a 15% chance of \$1,000,000	31% <sub>341</sub>	60% <sub>135</sub>	$p < 0.0001$
h	\$100 for sure or a 3% chance of \$7,000	8% <sub>139</sub>	21% <sub>70</sub>	$p < 0.01$
<i>Certain gains vs. Lower expected value gambles</i>		<i>Low</i>	<i>High</i>	
i	\$100 for sure or a 25% chance of \$200	7% <sub>68</sub>	10% <sub>20</sub>	n.s.
j	\$100 for sure or a 25% chance of \$300	14% <sub>137</sub>	18% <sub>39</sub>	n.s.
k	\$5 for sure or a 4% chance of \$80	29% <sub>84</sub>	36% <sub>50</sub>	n.s.
l	\$5 for sure or a 1% chance of \$80	27% <sub>37</sub>	37% <sub>38</sub>	n.s.
m	\$60 for sure or a 1% chance of \$5000	19% <sub>153</sub>	32% <sub>31</sub>	n.s.
<i>Certain losses vs. Lower expected value gambles</i>		<i>Low</i>	<i>High</i>	
n	Lose \$10 for sure or a 90% chance to lose \$50	24% <sub>29</sub>	6% <sub>16</sub>	n.s.
o	Lose \$100 for sure or a 75% chance to lose \$200	54% <sub>339</sub>	31% <sub>141</sub>	$p < 0.0001$
p	Lose \$100 for sure or a 50% chance to lose \$300	61% <sub>335</sub>	55% <sub>109</sub>	n.s.
q	Lose \$50 for sure or a 10% chance to lose \$800	44% <sub>180</sub>	23% <sub>56</sub>	$p < 0.01$
r	Lose \$100 for sure or a 3% chance to lose \$7000	63% <sub>68</sub>	28% <sub>57</sub>	$p < 0.0001$

*Table 3b*  
**The Reflection Effect for Low and High CRT Groups**

<i>Item</i>	<i>Percentage choosing gamble in the domain of gains and losses</i>	<i>CRT group</i>	
		<i>Low</i>	<i>High</i>
d	\$100 for sure or a 75% chance of \$200	19% <sub>843</sub>	38% <sub>475</sub>
o	Lose \$100 for sure or a 75% chance to lose \$200	54% <sub>339</sub>	31% <sub>141</sub>
h	\$100 for sure or a 3% chance of \$7,000	8% <sub>139</sub>	21% <sub>70</sub>
r	Lose \$100 for sure or a 3% chance to lose \$7000	63% <sub>68</sub>	28% <sub>57</sub>

Football League<sup>11</sup> and other employers to assess the intellectual abilities of their prospective hires; 944 completed an 18-item “need for cognition” scale (NFC), which measures the endorsement of statements like “the notion of thinking ab-

<sup>11</sup> Pat McNally, a Harvard graduate who later became a punter for the Cincinnati Bengals, was the only college football player to score a perfect 50 out of 50 on the Wonderlic—a score attained by only one person in 30,000. Of the 921 respondents who took it in these studies, the highest score was a 47.

Table 4  
Correlations Between Cognitive Measures

	<i>CRT</i>	<i>SAT</i>	<i>SAT<sub>M</sub></i>	<i>SAT<sub>V</sub></i>	<i>ACT</i>	<i>WPT</i>	<i>NFC</i>
<i>CRT</i>		434	434	434	667	921	944
<i>SAT</i>	.44		434	434	152	276	64
<i>SAT<sub>M</sub></i>	.46	.77		434	152	276	64
<i>SAT<sub>V</sub></i>	.24	.81	.28		152	276	64
<i>ACT</i>	.46	.77	.63	.67		466	190
<i>WPT</i>	.43	.49	.40	.37	.48		276
<i>NFC</i>	.22	.30	.21	.28	.30	.19	

stractly is appealing to me” (Cacioppo, Petty and Kao, 1984). Several hundred respondents also reported their scores on the Scholastic Achievement Test (SAT) or the American College Test (ACT), the two most common college entrance examinations.

Table 4 shows the correlations between cognitive measures. The numbers above the diagonal are the sample sizes from which these correlations were computed (the number of surveys that included both measures). For example, 152 respondents reported both SAT and ACT scores, and their correlation was 0.77. As expected, all measures correlate positively and significantly with one another. The moderate correlations suggest that all five tests likely reflect common factors, but may also measure distinct characteristics, as they purport to. I have proposed that the CRT measures “cognitive reflection”—the ability or disposition to resist reporting the response that first comes to mind. The need for cognition scale (NFC) is advanced as a measure of someone’s “tendency to engage in and enjoy thinking” (Cacioppo and Petty, 1982), but relies on self-reports rather than observed behavior. The Wonderlic Personnel Test (WPT) is intended to measure a person’s general cognitive ability, and the ACT and SAT are described as measures of academic “achievement.”

Although the various tests are intended to measure conceptually distinguishable traits, there are many likely sources of shared variance. For example, though the CRT is intended to measure cognitive reflection, performance is surely aided by reading comprehension and mathematical skills (which the ACT and SAT also measure). Similarly, though NFC and intelligence are distinguishable, the list of ways in which those with high NFC differ from those with low NFC (see Cacioppo et al., 1996) sounds very much like the list one would create if people were sorted on *any* measure of cognitive ability. Namely, those with higher NFC were found to do better on arithmetic problems, anagrams, trivia tests and college coursework, to be more knowledgeable, more influenced by the quality of an argument, to recall more of the information to which they are exposed, to generate more “task relevant thoughts” and to engage in greater “information-processing activity.”

The empirical and conceptual overlap between these tests suggests that they would all predict time and risk preferences and raises the question of their relative

Table 5  
**Correlations Between Cognitive Measures and Decision-Making Indices**

Cognitive measure	Choice under uncertainty (Preferences for gambles across domains)			
	Intertemporal choice	Gains		Losses
	Preference for patient option	Expected value favors gamble	Expected value favors sure gain	Expected value favors sure loss
CRT	+0.12**** <sub>3099</sub>	+0.22**** <sub>3150</sub>	+0.08** <sub>1014</sub>	-0.12**** <sub>1366</sub>
SAT	+0.07 <sub>387</sub>	+0.09 <sub>368</sub>	+0.07 <sub>149</sub>	-0.12* <sub>275</sub>
SAT <sub>M</sub>	-0.04 <sub>387</sub>	+0.19*** <sub>368</sub>	+0.05 <sub>149</sub>	-0.11 <sub>275</sub>
SAT <sub>V</sub>	+0.15** <sub>387</sub>	-0.03 <sub>368</sub>	+0.06 <sub>149</sub>	-0.08 <sub>275</sub>
ACT	+0.10* <sub>577</sub>	+0.14** <sub>549</sub>	+0.13* <sub>367</sub>	-0.01 <sub>358</sub>
WPT	+0.00 <sub>837</sub>	+0.13*** <sub>904</sub>	+0.08 <sub>287</sub>	-0.24**** <sub>546</sub>
NFC	+0.06 <sub>755</sub>	+0.13**** <sub>875</sub>	+0.03 <sub>497</sub>	-0.00 <sub>215</sub>

predictive validities. To assess this issue, I correlated the scores on the various cognitive measures with composite indices of decision-making characteristics formed from the time preference items in Table 2 or the risk preference items in Table 3. The composite scores registered the proportion of patient (or risk seeking) responses. For example, respondents might have been asked whether they prefer \$3,400 this month or \$3,800 next month, whether they would prefer a shorter massage in two weeks or a longer one in November and how much they would pay for overnight shipping of a book. Respondents who preferred the \$3800, the longer later massage and who were willing to pay less than the median person for express shipping would be coded as “patient” on all three items and would receive a score of 1. If they were patient on two of the three items, they would receive a score of 0.66, and so on. Thus, the indices are scores ranging from 0 to 1, in coarse or fine increments depending on how many questions the respondent answered.<sup>12</sup>

As shown in Table 5, the CRT was either the best or second-best predictor across all four decision-making domains and the only test related to them all. Thus,

<sup>12</sup> Composite indices were used to measure respondents’ general tendencies within a given decision-making domain and to permit aggregation across studies. However, unless respondents received identical items, their scores are not perfectly comparable. This issue is not vital for establishing the predictive validity of the CRT, because the correlations reflect the pattern plainly observable from the individual items. However, for the purpose of comparing the cognitive measures, composite indices are more problematic, because the full battery of cognitive tests was not typically given, and different studies involved different items. For example, at Carnegie Mellon University, respondents answered items b, d and l from Table 2 and items a and d from Table 3. The CRT was the only cognitive measure obtained for these respondents. Thus, these particular items will be disproportionately represented in the composite decision-making indices with which the CRT is correlated. This problem can be overcome by doing a pairwise comparison of cognitive measures only for those respondents who were given both. This more painstaking analysis generally confirms the implications of Table 5—namely, the different tests often function similarly, but the CRT is a bit more highly correlated with the characteristics of interest.

for researchers interested in separating people into cognitive groups, the CRT is an attractive test: it involves only three items and can be administered in a minute or two, yet its predictive validity equals or exceeds other cognitive tests that involve up to 215 items and take up to 3½ hours to complete (or which involve self-reports that cannot be readily verified).

## Sex Differences

Men scored significantly higher than women on the CRT, as shown in Table 6. The difference is not likely due to a biased sampling procedure, because there were no significant sex differences for any other cognitive measure, except SAT<sub>math</sub> scores, for which there was a modest difference corresponding to national averages. Nor can it be readily attributed to differences in the attention or effort expended on the survey, since women scored slightly *higher* on the Wonderlic test, which was given under identical circumstances (included as part of a 45-minute survey that recruited respondents were paid to complete).

It appears, instead, that these items measure something that men have more of. That something may be mathematical ability or interest, since the CRT items have mathematical content, and men generally score higher than women on math tests (Benbow and Stanley, 1980; Halpern, 1986; Hyde, Fennema and Lamon, 1990; Hedges and Nowell, 1995). However, men score higher than women on the CRT, even controlling for SAT math scores. Furthermore, even if one focuses only on respondents who gave the wrong answers, men and women differ. Women's mistakes tend to be of the intuitive variety, whereas men make a wider variety of errors. For example, the women who miss the "widgets" problem nearly always give the erroneous intuitive answer "100," whereas a modest fraction of the men give unexpected wrong answers, such as "20" or "500" or "1." For every CRT item (and several other similar items used in a longer variant of the test) the ratio of "intuitive" mistakes to "other" mistakes is higher for women than for men. Thus, the data suggest that men are more likely to reflect on their answers and less inclined to go with their intuitive responses.<sup>13</sup>

Because men score higher, the "high" CRT group is two-thirds men, whereas the "low" CRT group is two-thirds women. Thus, the differences between CRT groups may be revealing other male/female differences besides cognitive reflection. To remove this confound, Table 7 presents results split by both sex and CRT score for selected items, including a heretofore undiscussed item involving the willingness to pay for a coin flip in which "heads" pays \$100 and "tails" pays nothing.

Four facts are noteworthy. First, CRT scores are more highly correlated with time preferences for women than for men; the low and high groups differ more. Second, as suggested by most prior research (Byrnes, Miller and Schafer, 1999,

<sup>13</sup> One might draw the opposite conclusion from self-reports. Using the scale described earlier, respondents were asked "How long do you deliberate before reaching a conclusion?" Women reported *higher* scores than men (1.16 vs. 0.45;  $t_{186} = 2.32$ ;  $p < 0.05$ ).

**Table 6**  
**Sex Differences in Cognitive Measures**

<i>Test</i>	<i>Men</i>	<i>Women</i>	<i>Significance of group difference</i>
CRT	1.47	1.03	$p < 0.0001$
SAT	1334	1324	n.s.
SAT <sub>math</sub>	688	666	$p < 0.01$
SAT <sub>verbal</sub>	646	658	n.s.
ACT	26.7	26.3	n.s.
Wonderlic	26.2	26.5	n.s.
NFC	0.91	0.85	n.s.

present an overview), women were considerably more risk averse than men, and this remains true even after controlling for CRT score. Third, for the selected risk items, CRT is as important as sex. In other words, high-scoring women behave almost identically to low-scoring men (compare the upper left and lower right cells within each of the five items in the lower panel). Fourth, in contrast to the pattern observed for the time preference items, CRT scores are more highly correlated with risk preferences for men than for women.

The curious finding that CRT scores are more tightly linked with time preferences for women than for men, but are more tightly linked with risk preferences for men than for women held for the other tests of cognitive ability, as well. Expressed loosely, being smart makes women patient and makes men take more risks.<sup>14</sup> This result was unanticipated and suggests no obvious explanation. The only related finding of which I am aware is in a study by Shoda, Mischel and Peake (1990), who found that the patience of preschool girls was strongly related to their subsequent SAT scores, but the patience of preschool boys was not.

## Discussion

The instructions in studies of decision making commonly reassure respondents that “there are no right or wrong answers.” If this line is sincere, it implies that researchers will interpret such preferences as they would a choice between

<sup>14</sup> This conclusion can also be expressed less loosely. First, when faced with three mathematical reasoning problems (“bat and ball,” “widgets” and “lily pads”), certain responses that are plausibly construed as manifestations of intelligence (“5,” “5” and “47”) tend to correlate positively with certain other responses that are plausibly construed as expressions of patience (namely, an expressed willingness to wait for larger later rewards), and this tendency is more pronounced in women than men. Second, the production of the canonically correct responses tends also to correlate positively with certain responses that are plausibly construed as expressions of risk tolerance (namely, an expressed willingness to forego a smaller certain reward in favor of a probabilistic larger one), and this tendency is more pronounced in men than in women. Third, sex differences in risk seeking and in the degree of relation to CRT scores was true only in the domain of gains. For the selected loss items (n through r in Table 3), there were no sex differences.

*Table 7*  
**Results Split by Both CRT and Sex**  
*(percentage choosing patient option or mean response)*

<i>Intertemporal choice or judgment</i>	<i>Sex</i>	<i>CRT group</i>		<i>Significance of group difference</i>
		<i>Low</i>	<i>High</i>	
\$3400 this month or \$3800 next month	Men	39% <sub>170</sub>	60% <sub>84</sub>	$p < 0.01$
	Women	39% <sub>252</sub>	67% <sub>51</sub>	$p < 0.001$
\$100 this year or \$140 next year	Men	21% <sub>106</sub>	34% <sub>161</sub>	$p < 0.05$
	Women	25% <sub>194</sub>	49% <sub>70</sub>	$p < 0.001$
\$100 now or \$1100 in 10 years	Men	58% <sub>88</sub>	56% <sub>110</sub>	n.s.
	Women	43% <sub>186</sub>	57% <sub>68</sub>	$p < 0.05$
\$9 now or \$100 in 10 years	Men	40% <sub>123</sub>	43% <sub>178</sub>	n.s.
	Women	41% <sub>229</sub>	53% <sub>89</sub>	$p < 0.10$
Willingness to pay for overnight shipping of chosen book	Men	\$4.05 <sub>41</sub>	\$1.94 <sub>84</sub>	$p < 0.001$
	Women	\$4.54 <sub>95</sub>	\$2.19 <sub>40</sub>	$p < 0.001$

  

<i>Risky choice or judgment (percentage choosing risky option or mean response)</i>	<i>Sex</i>	<i>CRT group</i>		<i>Significance of group difference</i>
		<i>Low</i>	<i>High</i>	
\$100 for sure or a 75% chance of \$200	Men	26% <sub>239</sub>	43% <sub>244</sub>	$p < 0.0001$
	Women	16% <sub>398</sub>	29% <sub>130</sub>	$p < 0.01$
\$500 for sure or a 15% chance of \$1,000,000	Men	40% <sub>68</sub>	80% <sub>41</sub>	$p < 0.0001$
	Women	25% <sub>109</sub>	38% <sub>37</sub>	n.s.
\$1000 for sure or a 90% chance of \$5000	Men	59% <sub>103</sub>	81% <sub>151</sub>	$p < 0.001$
	Women	46% <sub>166</sub>	59% <sub>65</sub>	$p < 0.10$
\$100 for sure or a 3% chance of \$7000	Men	6% <sub>36</sub>	30% <sub>44</sub>	$p < 0.01$
	Women	8% <sub>99</sub>	8% <sub>24</sub>	n.s.
Willingness to pay for a coin flip, where "HEADS" pays \$100 and "TAILS" pays nothing.	Men	\$13.00 <sub>54</sub>	\$20.00 <sub>59</sub>	$p < 0.001$
	Women	\$11.00 <sub>12</sub>	\$12.00 <sub>36</sub>	n.s.

apples and oranges—as a primitive that neither requires nor permits further scrutiny.

However, unlike a preference between apples and oranges, time and risk preferences are sometimes tied so strongly to measures of cognitive ability that they effectively function as such a measure themselves.<sup>15</sup> For example, when a choice

<sup>15</sup> To encourage respondents to consider each choice carefully, and independently from the other items, several “filler” choices were inserted between the “focal items.” An analysis of these responses shows that CRT scores are unrelated to preferences between apples and oranges, Pepsi and Coke, beer



between a sure \$500 and a 15 percent chance of \$1,000,000 was presented to respondents along with an eight-item version of the CRT, only 25 percent of those who missed all eight problems chose the gamble, compared to 82 percent among those who solved them all. Should this result be interpreted to mean that choosing the gamble is the “correct” response for this item?

The position that some preferences are better than others and that cognitive ability is one indicator of the “better” preference is not unprecedented. Savage (1954) argued that increased understanding ought to increase the frequency of the “truly” normative response; that preferences that initially contradict some normative principle may not survive thorough deliberation (what he termed “reflective equilibrium”).<sup>16</sup> Stanovich and West (2000) extended these views, by arguing that increased understanding may arise from superior intellect (as well as from extended deliberation or reflection or instruction). In response to those contending that judgments commonly labeled as errors or biases are actually equally good answers to different interpretations of the question (for example, Hilton, 1995), Stanovich and West argued that if smarter respondents were more likely to give canonically correct answers, the other answers must not be equally good after all.<sup>17</sup>

Some, however, reject the notion that a correlation between (some measure of) cognitive ability and some particular response identifies the “better” response. For example, Sternberg (2000, pp. 697–698) argues: “[T]o characterize people with high SAT scores as those who should set the norm for what is somehow true or right seems to be off target. People with high SAT scores have high levels of certain kinds of cognitive abilities. They have no monopoly on quality of thinking and certainly no monopoly on truth.”

The prevalence of this view could be directly tested. Respondents could be shown the respective test scores of those who chose the sure \$500 and those who chose the 15 percent chance of \$1,000,000. If Sternberg’s view is widely shared, this manipulation would have no effect. If, on the other hand, the correlation between cognitive ability and preference held normative force, making respondents aware of it would cause many of them to choose the gamble.

Of course, the weight that should be placed on the opinions of those with higher cognitive abilities clearly depends on the type of decision in question. If one were deciding between a fixed- and variable-interest mortgage, imitating one’s

and wine or rap concerts and ballet. However, CRT scores are strongly predictive of the choice between *People* magazine and the *New Yorker*. Among the low CRT group, 67 percent preferred *People*. Among the high CRT group, 64 percent preferred the *New Yorker*.

<sup>16</sup> Slovic and Tversky (1974) use an eloquent and entertaining mock debate between Allais and Savage to illustrate opposing views on the related issue of whether the opinions of people who have deliberated longer over an issue ought to count more.

<sup>17</sup> Along similar lines, Bar Hillel (1991, p. 413) comments: “Many writers have attempted to defend seemingly erroneous responses by offering interpretations of subjects’ reasoning that rationalizes their responses. Sometimes, however, this charitable approach has been misguided, either because the subjects are quick to acknowledge their error themselves once it is pointed out to them, or because the interpretation required to justify the response is even more embarrassing than the error it seeks to excuse.”

brilliant neighbor seems prudent. However, if one were deciding between an apple and an orange, Einstein's preference for apples seems irrelevant.

Thus, a relation between cognitive ability and preference does not, by itself, establish the correct choice for any particular individual. Two individuals with different cognitive abilities may experience outcomes differently, which may warrant different choices (for example, what magazines to read or movies to attend). But with respect to the example motivating this discussion, one must ask whether it is plausible that people of differing cognitive abilities experience increments of wealth as differently as their choices suggest. It seems exceedingly unlikely that the low CRT group has a marked kink in their utility function around  $\$W + 500$ , beyond which an extra  $\$999,500$  confers little additional benefit. It seems more reasonable, instead, to override the conventional caveat about arguing with tastes (Becker and Stigler, 1977) and conclude that choosing the  $\$500$  is the "wrong answer"—much as 10 cents is the wrong answer in the "bat and ball" problem.

Whatever stance one adopts on the contentious normative issues of whether a preference can be "wrong" and whether more reflective people make "better" choices, respondents who score differently on the CRT make *different* choices, and this demands *some* explanation.

■ *I thank Dan Ariely, Scott Armstrong, Daniel Benjamin, Brett Boshco, Eric Bradlow, Craig Fox, Kerri Frederick, Steve Garcia, Timothy Heath, James Hines, Eric Johnson, Daniel Kahneman, Robyn LeBoeuf, George Loewenstein, Leif Nelson, Nathan Novemsky, Greg Pogarsky, Drazen Prelec, Daniel Read, Eldar Shafir, Timothy Taylor, Catherine Tucker, Michael Waldman and Jaclyn Zires for comments received on earlier drafts. A special thanks to Steve Garcia, who coordinated most of the surveys generating the data summarized here. As always (but particularly in this case), the views expressed or implied are those of the author alone.*

## References

- Bar-Hillel, Maya. 1991. "Commentary on Wolford, Taylor, and Beck: The Conjunction Fallacy?" *Memory and Cognition*. 19:4, pp. 412–14.
- Becker, Gary and George Stigler. 1977. "De Gustibus Non est Disputandum." *American Economic Review*. 67:2, pp. 76–90.
- Benbow, Camilla P. and J. C. Stanley. 1980. "Sex Differences in Mathematical Ability: Fact or Artifact?" *Science*. 210:4475, pp. 1262–264.
- Benjamin, Daniel J. and Jesse M. Shapiro. 2005. "Who is 'Behavioral?' Cognitive Ability and Anomalous Preferences." Working paper, Harvard University.
- Byrnes, James P., David C. Miller and William D. Schafer. 1999. "Gender Differences in Risk Taking: A Meta-Analysis." *Psychological Bulletin*. 125:3, pp. 367–83.
- Cacioppo, John T. and Richard E. Petty. 1982. "The Need for Cognition." *Journal of Personality and Social Psychology*. 42:1, pp. 116–31.
- Cacioppo, John T., Richard E. Petty and Chuan Feng Kao. 1984. "The Efficient Assessment of Need for Cognition." *Journal of Personality Assessment*. 48:3, pp. 306–07.
- Cacioppo, John T., Richard E. Petty, Jeffrey A. Feinstein and W. Blair G. Jarvis. 1996. "Dis-

- positional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition." *Psychological Bulletin*. 119:2, pp. 197–253.
- Chaiken, Shelly and Yaacov Trope.** 1999. *Dual-Process Theories in Social Psychology*. New York: Guilford Press.
- Donkers, Bas, Bertrand Melenberg and Arthur van Soest.** 2001. "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach." *Journal of Risk and Uncertainty*. 22:2, pp. 165–95.
- Epstein, Seymour.** 1994. "Integration of the Cognitive and Psychodynamic Unconscious." *American Psychologist*. 49:8, pp. 709–24.
- Frederick, Shane.** 2002. "Automated Choice Heuristics," in *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich, D. Griffin and D. Kahneman, eds. New York: Cambridge University Press, pp. 548–58.
- Frederick, Shane, George Loewenstein and Ted O'Donoghue.** 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*. 40:2, pp. 351–401.
- Funder, David C. and Jack Block.** 1989. "The Role of Ego-Control, Ego-Resiliency, and IQ in Delay of Gratification in Adolescence." *Journal of Personality and Social Psychology*. 57:6, pp. 1041–050.
- Halpern, Diane F.** 1986. *Sex Differences in Cognitive Abilities*. Hillsdale, N.J.: Erlbaum.
- Hedges, Larry V. and Amy Nowell.** 1995. "Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals." *Science*. July 7, 269, pp. 41–45.
- Hilton, Denis J.** 1995. "The Social Context of Reasoning: Conversational Inference and Rational Judgment." *Psychological Bulletin*. September, 118, pp. 248–71.
- Hyde, Janet Shibley, Elizabeth Fennema and Susan J. Lamon.** 1990. "Gender Differences in Mathematics Performance: A Meta-Analysis." *Psychological Bulletin*. 107:2, pp. 139–55.
- Jensen, Arthur R.** 1998. *The g Factor: The Science of Mental Ability*. Westport, Conn.: Praeger.
- Kahneman, Daniel and Shane Frederick.** 2002. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich, D. Griffin and D. Kahneman, eds. New York: Cambridge University Press, pp. 49–81.
- Kahneman, Daniel and Amos Tversky.** 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*. 47:2, pp. 263–91.
- Kirby, Kris N., Gordon C. Winston and Mariana Sentisteban.** 2005. "Impatience and Grades: Delay-Discount Rates Correlate Negatively with College GPA." *Learning and Individual Differences*. Forthcoming.
- Lubinski, David and Lloyd Humphreys.** 1997. "Incorporating General Intelligence into Epidemiology and the Social Sciences." *Intelligence*. 24:1, pp. 159–201.
- Melikian, Levon.** 1959. "Preference for Delayed Reinforcement: An Experimental Study among Palestinian Arab Refugee Children." *Journal of Social Psychology*. 50, pp. 81–86.
- Mischel, Walter.** 1974. "Processes in Delay of Gratification," in *Advances in Experimental Social Psychology*. L. Berkowitz, ed. San Diego, Calif.: Academic Press, pp. 249–92.
- Monterosso, John, Ronald Ehrman, Kimberly L. Napier, Charles P. O'Brien and Anna Rose Childress.** 2001. "Three Decision-Making Tasks in Cocaine-Dependent Patients: Do They Measure the Same Construct?" *Addiction*. 96:12, pp. 1825–837.
- Nagin, Daniel S. and Greg Pogarsky.** 2003. "An Experimental Investigation of Deterrence: Cheating, Self-Serving Bias, and Impulsivity." *Criminology*. 41:1, pp. 501–27.
- Parker, Andrew M. and Baruch Fischhoff.** 2005. "Decision-Making Competence: External Validation through an Individual-Differences Approach." *Journal of Behavioral Decision Making*. 18:1, pp. 1–27.
- Rabin, Matthew.** 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*. 68:5, pp. 1281–292.
- Rae, John.** 1834. *The New Principles of Political Economy*. Reprinted in 1905 as *The Sociological Theory of Capital*. New York: Macmillan.
- Savage, Leonard J.** 1954. *The Foundations of Statistics*. New York: Wiley.
- Shoda, Yuichi, Walter Mischel and Philip K. Peake.** 1990. "Predicting Adolescent Cognitive and Self-Regulatory Competencies from Preschool Delay of Gratification: Identifying Diagnostic Conditions." *Developmental Psychology*. 26:6, pp. 978–86.
- Sloman, Steven A.** 1996. "The Empirical Case for Two Systems of Reasoning." *Psychological Bulletin*. 119:1, pp. 3–22.
- Slovic, Paul and Amos Tversky.** 1974. "Who Accepts Savage's Axiom?" *Behavioral Science*. 19:4, pp. 368–73.
- Stanovich, Keith E. and Richard F. West.** 2000. "Individual Differences in Reasoning: Implications for the Rationality Debate?" *Behavioral and Brain Sciences*. 22:5, pp. 645–726.
- Sternberg, Robert J.** 2000. "The Ability is not General, and Neither are the Conclusions. [Response to K. E. Stanovich and R.F. West.]" *Behavioral and Brain Sciences*. 23:5, pp. 697–98.