On the Ball: Cognitive Reflection and Decision Making

Shane Frederick

Shane Frederick is Assistant Professor of Management Science, Sloan School of Management, Massachusetts Institute of Technology, Cambridge Massachusetts.

Email: shanefre@mit.edu

People 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 or policy. 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 a fairly homogenous bunch. Third, characterizing performance differences on cognitive tasks requires terms ("IQ" and "aptitudes" and such) that are often objected to because of their association with discriminatory policies. In short, researchers may be reluctant to study something they don't find interesting, that isn't 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. Thus, to provoke interest in this neglected topic, this paper introduces a three-item "Cognitive Reflection Test" (CRT), as a simple measure of one type of cognitive ability. I'll show that CRT scores are highly predictive of the very types of choices that feature prominently in tests of decision making theories, like expected utility theory and prospect theory. Indeed, test scores are sometimes so diagnostic of preferences 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. Some of these cognitive measures exhibit differences between men and women and I discuss how these relate to sex differences in time and risk preferences. The final section broaches the issue of how one should interpret 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 (Sloman, 1996; Chaiken & 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.[1]

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.

[1] For a discussion of the distinction between System 1 and System 2 in the context of choice heuristics, see Frederick (2002).

By contrast, consider the problem below:

> 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 recognize that the difference between $1 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." 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.[2]

---

[2] 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 answer dominates. Second, even among those responding correctly, the wrong answer was often considered first, as is apparent from introspection, verbal reports, and from the fact that 10 cents was often crossed out next to 5 cents (but never the other way around). Third, in one study, respondents not only attempted to solve the problems themselves, but also estimated the proportion of *other* respondents who would correctly solve them. Respondents who missed the problems thought they were easier than those who solved them. For example, in the "bat & ball" problem, those who said "10 cents" estimated that 92% of people would correctly solve it, whereas those who said "5 cents" estimated that "only" 62% would. (Both were considerable overestimates.) Thus, errors occurred because the problems seemed too easy to miss, not because respondents failed to comprehend what was being asked. Fourth, respondents do much better if their computational machinery is not highjacked by an aggressively intrusive intuitive answer. The following "banana & bagel" problem is solved about twice as often as the "bat & ball" problem. ("*A banana and a bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost?*")

Figure 1. The Cognitive Reflection Test (CRT)

(1) A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?

____ cents

(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

Over a 26-month period beginning in January, 2003, the CRT was administered to 3,428 respondents in 35 separate studies that also measured various decision making characteristics, like time and risk preferences. 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.[3] 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 percent 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.[4]

[3] There were three exceptions to this: (1) the participants from Carnegie Mellon University completed the survey as part of class; (2) the 4th 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.

[4] Most of the respondents were college students from highly selective or "elite" institutions. The mean reported SAT score (1331) exceeds the national average (1026) by almost three standard deviations. The mean ACT and WPT scores were similarly elevated. Thus, even the two "extreme" groups that formed the basis for most statistical comparisons were more similar in cognitive abilities than two respondents randomly selected 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.

Table 1. CRT scores, by location

Locations at which data were collected [5]	mean CRT score	Percent scoring 0, 1, 2, or 3				N=
		0 "low"	1	2	3 "high"	
Massachusetts Institute of Technology	2.18	7%	16%	30%	48%	61
Princeton University	1.63	18%	27%	28%	26%	121
Boston fireworks display*	1.53	24%	24%	26%	26%	195
Carnegie Mellon University	1.51	25%	25%	25%	25%	746
Harvard University**	1.43	20%	37%	24%	20%	51
University of Michigan: Ann Arbor	1.18	31%	33%	23%	14%	1267
Web-based studies***	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**

Cognitive Reflection and Time Preferences

The notion that more intelligent people are more patient – that they devalue or "discount" future rewards less steeply – has prevailed for some time. For example, in the *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 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. He found that the children who opted for the promissory note scored slightly, but significantly, higher on an

[5]* Respondents in this study were people picnicking along the banks of the Charles River prior to the July 4th 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 along 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.)

** This was the only location at which the number of male and female students was not approximately equal. Of the fifty one participants in this study, forty two were women.

*** These were participants in two online studies. Respondents consisted primarily of students from universities in the northeast and midwest.

intelligence test based on an assessment of those drawings.[6] Funder and Block (1989) paid fourteen-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 IQ's 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 (see 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 eighteen, to prefer a larger later reward over a smaller sooner one (e.g. \$120 in four weeks to \$100 tomorrow). Using small real rewards, Benjamin and Shapiro (2005) found that respondents with higher SAT math scores (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.) Monterrosso 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 GPAs correlated positively with those bids).

Collectively, these studies offer some support for the view that cognitive ability and time preference are somehow connected, though none has identified the types of intertemporal decisions over which cognitive ability exerts influence, nor explained why it does so.[7] Toward this end, I examined the relation between CRT scores and a wide variety of items intended to measure various aspects of "time preference." As shown in Table 2, these included several hypothetical choices between an immediate reward and a

[6] The children in this study ranged from 5 to 12 years. Given this wide range, 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 general cognitive ability.

[7] This is not quite true. 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 (e.g. "think fun") was suggested. In the other three conditions, patient behavior was actually *negatively* correlated with subsequent SAT scores.

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 between a smaller immediate loss and a larger delayed loss (items j and k).[8] Item l asked respondents to state their maximum willingness to pay to have a book shipped overnight rather than waiting two weeks.[9] Item m involved a real payment. Respondents specified the smallest amount of money in four days that they would prefer to $170 in two months, knowing that one of them would be selected to actually receive their choice. 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).[10]

Table 2 shows the responses of the low and high CRT groups for each of the sixteen items. The reported value is either the percent 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 χ^2test (for dichotomous responses) or a t-test (for continuous responses).

[8] 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 disutilty 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.

[9] Respondents were asked to indicate their favorite book from a small set and to specify how much they would pay for overnight shipping rather than waiting two weeks. The free book was described as a reward for participating in an on-line survey for Amazon.com. The set of books included: *Longitudes and Attitudes: Exploring the World After September 11*, *Guns, Germs & Steel: Exploring the Fate of Human Societies*, *Against Love: A Polemic*, *The Selfish Gene*, and the horror novel *Thinner*.

[10] Among the items in Table 2, men were more patient for items c, k, l, and 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

item	Intertemporal choice or judgment (percent choosing patient option or mean response)	CRT group		Signif of stat. diff.
		low	high	
a	\$3400 this month or \$3800 next month	**35%** $_{611}$	**60%** $_{196}$	**p<0.0001**
b	\$100 now or \$140 next year	**22%** $_{409}$	**37%** $_{297}$	**p<0.0001**
c	\$100 now or \$1100 in 10 years	**47%** $_{283}$	**57%** $_{208}$	**p<0.05**
d	\$9 now or \$100 in 10 years	**40%** $_{364}$	**46%** $_{277}$	**p<0.10**
e	\$40 immediately or \$1000 in 10 years	**50%** $_{135}$	**59%** $_{83}$	**n.s.**
f	\$100 now or \$20 every year for 7 years	**28%** $_{60}$	**43%** $_{28}$	**n.s.**
g	\$400 now or \$100 every year for 10 years	**64%** $_{44}$	**72%** $_{43}$	**n.s.**
h	\$1000 now or \$100 every year for 25 years	**52%** $_{295}$	**49%** $_{99}$	**n.s.**
i	30 min. massage in 2 weeks or 45 min. massage in Nov.	**28%** $_{272}$	**27%** $_{126}$	**n.s.**
j	lose \$1000 this year or lose \$2000 next year	**78%** $_{166}$	**73%** $_{86}$	**n.s.**
k	tooth pulled today or tooth pulled in 2 weeks	**59%** $_{430}$	**65%** $_{242}$	**n.s.**
l	willingness to pay for overnight shipping of chosen book	**\$4.54** $_{150}$	**\$2.18** $_{163}$	**p<0.0001**
m	smallest amount in 4 days preferred to \$170 in 2 months	**\$117** $_{84}$	**\$131** $_{50}$	**p<0.05**
n	How impulsive are you?	**+1.01** $_{110}$	**-0.21** $_{47}$	**p<0.001**
o	How much do you tend to procrastinate?	**+1.05** $_{110}$	**+1.06** $_{47}$	**n.s.**
p	How much do you think about your future?	**+2.49** $_{110}$	**+1.64** $_{47}$	**p<0.01**
q	How much do you worry about inflation?	**-1.16** $_{110}$	**+0.11** $_{47}$	**p<0.01**

As predicted, 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 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 – 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 amount is interpreted as being denominated in nominal units).[11] 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%), exceeds market interest rates. However, such reasons do not apply with the same force for short term options; they are not sufficiently compelling to justify choosing \$3400 this month over \$3800 next month (which implies an annual discount rate of

[11] Frederick, Loewenstein, & 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.

280%). Hence, for choices like items (a) and (b), where the careful deliberation associated with "System 2" ought to strongly oppose one's intuitive "System 1" preference for the more immediate reward (see McClure et al., 2004), one observes considerable differences between CRT groups. Conversely, the differences between CRT groups are negligible for many of the other items, in which additional reflection would not make such a strong case against the more immediate reward.

Thus, greater cognitive reflection seemingly fosters the recognition or appreciation of considerations (like interest rates) that may favor the later larger reward. Whether it also influence other determinants of intertemporal choices (like *pure* time preference) remains unclear. CRT scores were unrelated to preferences for the massage and toothpull items which were intended as measures of pure time preference, because interest rates should play no role.[12] However, those in the low CRT group (the "cognitively impulsive") were willing to pay significantly more for the overnight shipping of a chosen book (item l), which <u>does</u> seem like an expression 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.

[12] Preferring the shorter more immediate massage is arguably irrational, because respondents cannot invest it, will still be alive to enjoy the delayed massage, still be stressed and sore, still like massages, and still derive greater benefits from longer ones.

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 empirical research. 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% chance of 45 florins over 30 florins for sure).[13] Benjamin and Shapiro (2005) found that students with higher scores on 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). I am unaware of any other relevant studies.

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 3. 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 3 are aggregated, the high CRT group gambled significantly more often than the low CRT group (31 percent vs. 19 percent; χ^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 adopt that as the choice criterion.[14] For items involving losses (lower panel), the higher CRT group was *less* risk seeking; they were more willing accept a sure loss to avoid playing a gamble with lower (more negative) expected value.

---

[13] At the time, a Dutch florin (Dfl) was worth about 50 cents.

[14] As expected, the gamble was not popular among *either* group for *any* of the items i through m, since risk aversion and expected value both militate against it. However, any factors favoring the gamble over the sure thing (e.g. valuing the excitement 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 (i.e., the high CRT group, as judged from items a through h). These particular "anti-EV" gambles were designed, in part, to have some chance of being chosen (the sure amounts were small and the expected values were typically close). Including choices in which the gambles lacked these properties (e.g. offering a choice between $4,000 for sure and a 50% 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.

Table 3. Risk seeking behavior among low and high CRT groups

	Percent choosing riskier option	CRT group		Stat. signif.
item	**certain gains vs. *higher* EV gambles**	**low**	**high**	
a	$1,000 for sure or a 90% chance of $5,000	**52%**$_{280}$	**74%**$_{225}$	**p<0.0001**
b	$100 for sure or a 90% chance of $500	**56%**$_{95}$	**78%**$_{92}$	**p<0.01**
c	$1,000 for sure or a 75% chance of $4,000	**37%**$_{264}$	**57%**$_{102}$	**p<0.001**
d	$100 for sure or a 75% chance of $200	**19%**$_{843}$	**38%**$_{475}$	**p<0.0001**
e	$100 for sure or a 75% chance of $150	**10%**$_{217}$	**34%**$_{94}$	**p<0.0001**
f	$100 for sure or a 50% chance of $300	**47%**$_{68}$	**75%**$_{20}$	**p<0.05**
g	$500 for sure or a 15% chance of $1,000,000	**31%**$_{341}$	**60%**$_{135}$	**p<0.0001**
h	$100 for sure or a 3% chance of $7,000	**8%**$_{139}$	**21%**$_{70}$	**p<0.01**
	certain gains vs. *lower* EV gambles	**low**	**high**	
i	$100 for sure or a 25% chance of $200	**7%**$_{68}$	**10%**$_{20}$	**n.s.**
j	$100 for sure or a 25% chance of $300	**14%**$_{137}$	**18%**$_{39}$	**n.s.**
k	$5 for sure or a 4% chance of $80	**29%**$_{84}$	**36%**$_{50}$	**n.s.**
l	$5 for sure or a 1% chance of $80	**27%**$_{37}$	**37%**$_{38}$	**n.s.**
m	$60 for sure or a 1% chance of $5000	**19%**$_{153}$	**32%**$_{31}$	**p<0.10**
	certain losses vs. *lower* EV gambles	**low**	**high**	
n	lose $10 for sure or a 90% chance to lose $50	**24%**$_{29}$	**6%**$_{16}$	**n.s.**
o	lose $100 for sure or a 75% chance to lose $200	**54%**$_{339}$	**31%**$_{141}$	**p<0.0001**
p	lose $100 for sure or a 50% chance to lose $300	**61%**$_{335}$	**55%**$_{109}$	**n.s.**
q	lose $50 for sure or a 10% chance to lose $800	**44%**$_{180}$	**23%**$_{56}$	**p<0.01**
r	lose $100 for sure or a 3% chance to lose $7000	**63%**$_{68}$	**28%**$_{57}$	**p<0.0001**

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 & Tversky, 1979). As shown in Table 3B, 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. However, there is no reflection effect among the high CRT group.

Table 3B The reflection effect for low and high CRT groups

item	Percent choosing gamble in the domain of gains and losses	CRT group	
		low	**high**
d	$100 for sure or a 75% chance of $200	**19%**$_{843}$	**38%**$_{475}$
o	lose $100 for sure or a 75% chance to lose $200	**54%**$_{339}$	**31%**$_{141}$
	χ^2 test	**p<0.0001**	**n.s.**
h	$100 for sure or a 3% chance of $7,000	**8%**$_{139}$	**21%**$_{70}$
r	lose $100 for sure or a 3% chance to lose $7000	**63%**$_{68}$	**28%**$_{57}$
	χ^2 test	**p<0.0001**	**n.s.**

This result starkly shows the importance of considering cognitive ability when evaluating the descriptive validity of a theory of decision making. For these items, low CRT respondents behave exactly as prospect theory predicts.[15] However, the descriptively accuracy of expected utility theory improves for high CRT respondents, who are as apt to gamble in the domain of gains as in the domain of losses.[16]

Is the CRT just another IQ test?

Of the 3428 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 National Football League[17] 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 idea of thinking abstractly 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

---

[15] A subset of respondents took an eight item test, which included the three CRT items. The contrast between items d and o is even more evident for the extreme scorers on this longer test. Among those scoring 0 out of 8, gambling for losses was three times more popular than gambling for gains. Among those scoring 8 out of 8, gambling for gains was twice as popular as gambling for losses.

[16] Although the descriptive accuracy of expected utility theory markedly *improves* for respondents with higher scores, it cannot explain why a 75% 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; see Rabin, 2000).

[17] Pat McInally, a Harvard graduate who later became a punter for the Cincinatti Bengals, was the only college football player to score a perfect 50 out of 50 on the Wonderlic – a score attained by only 1 person in 30,000. Of the 921 respondents who took it in these studies, the highest score was a 47.

"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."

Table 4. Correlations between cognitive measures.

	CRT	**SAT**	**SAT_M**	**SAT_V**	**ACT**	**WPT**	**NFC**
CRT		434	434	434	667	921	944
SAT	.44		434	434	152	276	64
SAT_M	.46	.77		434	152	276	64
SAT_V	.24	.81	.28		152	276	64
ACT	.46	.77	.63	.67		466	190
WPT	.43	.49	.40	.37	.48		276
NFC	.22	.30	.21	.28	.30	.19	

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 on it is surely aided by reading comprehension and mathematical skills (which the ACT and SAT also measure).[18] Similarly, though Cacioppo et al. (1996, p. 239) claim that NFC is "clearly separable" from intelligence, their list of ways in which those with high NFC were found to differ from those with low NFC 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 predictive validities. To assess these things, I correlated the scores on the various cognitive measures with composite indices of decision making characteristics formed

[18] Conversely, the Princeton Review Board could make the SAT into a "giant CRT" by selecting items that yield an incorrect intuitive answer, and then including that answer as one of the response options. However, a cursory review of recent SAT prep manuals suggests that they rarely do so. The following item, which evokes no intuitive answer, is more typical:

The average of 8, 13, x and y is 6. The average of 15, 9, x, and x is 8. What is the value of y?

(A) -1 (B) 0 (C) 4 (D) 6 (E) 8

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, a respondent might have been asked whether they prefer $3400 this month or $3800 next month, how much they would pay for overnight shipping of a book, and whether they would prefer a shorter massage in 2 weeks or a longer one in November. If they preferred the $3800, the longer later massage, and were willing to pay less than the median person for express shipping, they 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, all of the indices are scores ranging from 0 to 1, in coarse or fine increments depending on how many questions the repondent answered.[19]

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, for researchers interested in separating people into cognitive groups, the CRT is an attractive test: it involves only 3 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 3 ½ hours to complete (or which involve self reports that cannot be readily verified).

[19] Composite indices were used to measure respondents' general tendencies within a given decision making domain and 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.

Table 5. Correlations between cognitive measures and decision making indices.

	INTER-TEMPORAL CHOICE	CHOICE UNDER UNCERTAINTY (Preferences for Gambles Across Domains)		
		GAINS		LOSSES
Cognitive Measure	Preference for patient option	EV favors gamble	EV favors sure gain	EV favors sure loss
CRT	+0.12 **** 3099	+0.22**** 3150	+0.08** 1014	- 0.12**** 1366
SAT	+0.07 387	+0.09 368	+0.07 149	- 0.12* 275
SAT_M	- 0.04 387	+0.19*** 368	+0.05 149	- 0.11 275
SAT_V	+0.15** 387	- 0.03 368	+0.06 149	- 0.08 275
ACT	+0.10* 577	+0.14** 549	+0.13* 367	- 0.01 358
WPT	+0.00 837	+0.13*** 904	+0.08 287	- 0.24**** 546
NFC	+0.06 755	+0.13**** 875	+0.03 497	- 0.00 215

Sex Differences

As shown in Table 6, men scored significantly higher than women on the CRT. 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_{math} 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 forty five minute survey that recruited respondents were paid to complete).

Table 6. Sex differences in cognitive measures

Test	Men	Women	Significance of group difference
CRT	1.47	1.03	$p<0.0001$
SAT	1334	1324	n.s.
SAT_{math}	688	666	$p<0.01$
SAT_{verbal}	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.

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). However, men score higher than women on the CRT, even controlling for SAT math scores. Furthermore, even if one focuses only on respondents who give 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 the men who miss it more frequently give other 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.[20]

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.

[20] 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 7

Intertemporal choice or judgment (percent choosing patient option or mean response)	SEX	CRT		Stat. signif.
		low	**high**	
$3400 this month or $3800 next month	**men**	**39%**$_{170}$	**60%**$_{84}$	**p<0.01**
	women	**39%**$_{252}$	**67%**$_{51}$	**p<0.001**
$100 this year or $140 next year	**men**	**21%**$_{106}$	**34%**$_{161}$	**p<0.05**
	women	**25%**$_{194}$	**49%**$_{70}$	**p<0.001**
$100 now or $1100 in 10 years	**men**	**58%**$_{88}$	**56%**$_{110}$	**n.s.**
	women	**43%**$_{186}$	**57%**$_{68}$	**p<0.05**
$9 now or $100 in 10 years	**men**	**40%**$_{123}$	**43%**$_{178}$	**n.s.**
	women	**41%**$_{229}$	**53%**$_{89}$	**p<0.10**
willingness to pay for overnight shipping of chosen book	**men**	**$4.05**$_{41}$	**$1.94**$_{84}$	**p<0.001**
	women	**$4.54**$_{95}$	**$2.19**$_{40}$	**p<0.001**
Risky choice or judgment (percent choosing risky option or mean response)	**SEX**	**CRT**		**Stat. signif.**
		low	**high**	
$100 for sure or a 75% chance of $200	**men**	**26%**$_{239}$	**43%**$_{244}$	**p<0.0001**
	women	**16%**$_{398}$	**29%**$_{130}$	**p<0.01**
$500 for sure or a 15% chance of $1,000,000	**men**	**40%**$_{68}$	**80%**$_{41}$	**p<0.0001**
	women	**25%**$_{109}$	**38%**$_{37}$	**n.s.**
$1000 for sure or a 90% chance of $5000	**men**	**59%**$_{103}$	**81%**$_{151}$	**p<0.001**
	women	**46%**$_{166}$	**59%**$_{65}$	**p<0.10**
$100 for sure or a 3% chance of $7000	**men**	**6%**$_{36}$	**30%**$_{44}$	**p<0.01**
	women	**8%**$_{99}$	**8%**$_{24}$	**n.s.**
willingness to pay for a coin flip, where "HEADS" pays $100 and "TAILS" pays nothing.	**men**	**$13**$_{54}$	**$20**$_{59}$	**p<0.001**
	women	**$11**$_{102}$	**$12**$_{36}$	**n.s.**

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 (see Byrnes, Miller, & Schafer, 1999 for 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, the differences between CRT groups are about the same as the differences between sexes. In other words, low scoring men behave almost identically to high scoring women. (Compare the upper left and lower right cells within each of the five items in the lower panel). By contrast, high scoring men are considerably more risk tolerant than low scoring women. (Compare the lower left and upper right cells). 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 test 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 test of cognitive ability as well. Expressed loosely, being smart makes women patient, and makes men more risk seeking.[21] 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 SAT scores obtained over a decade later, but the patience of preschool boys was not.

Discussion

In studies of decision making, the instructions commonly reassure respondents that "there are no right or wrong answers." If sincere, this line implies that researchers will interpret all preferences as they would a choice between 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 sometime tied so strongly to measures of cognitive ability that they effectively function as such a measure themselves.[22] For example, when a choice between a sure $500 and a 15% chance of $1,000,000 was presented to respondents (along with an eight item version of the CRT), only 25% of those who missed all eight problems chose the gamble, compared to 82% among those who solved them all.[23]

---

[21] This can also be expressed less loosely. First, when faced with three mathematical reasoning problems ("bat & ball," "widgets," and "lilypads"), 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 a risk tolerance (namely, an expressed willingness to forego a smaller certain reward in favor of a probabalistic 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), women and men were essentially indistinguishable.

[22] 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 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% preferred *People*. Among the high CRT group, 64% preferred *The New Yorker*.

[23] Other differences were less spectacular, but still striking. In a choice between $3400 this month and $3800 next month, the larger delayed reward was chosen only 33% of the "0s" but by 74% of the "8s". In a choice between $1000 for sure and a 75% chance of $4000, the gamble was chosen by 24% of the "0s" but by 60% of the "8s." And so on.

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 which initially contradict some normative principle may not survive thorough deliberation (what he termed "reflective equilibrium").[24] 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.[25]

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: "… to 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."

I doubt that Sternberg's agnosticism is widely shared. I suspect that if respondents were shown the respective test scores of those who chose the sure $500 vs. those who chose the 15% chance of $1,000,000, they would, in fact, feel more disposed to take the gamble; the correlation between cognitive ability and preference would hold some normative force for them.

Clearly, though, the weight that should be placed on the opinions of those with higher abilities depends on the type of ability and the type of decision in question. If a

---

[24] 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.

[25] 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."

person were deciding which piece to move in chess and were told what Gary Kasparov would do, it seems advisable to do the same thing. If, however, one were deciding between an apple or 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 really 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 & 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.

Acknowledgements

I thank Dan Ariely, Scott Armstrong, Craig Fox, Kerri Frederick, Steve Garcia, Tim Heath, James Hines, Eric Johnson, Daniel Kahneman, Robyn LeBoeuf, George Loewenstein, Leif Nelson, Nathan Novemsky, Greg Pogarsky, Drazen Prelec, Daniel Read, Eldar Shafir, Tim Taylor, Michael Waldman, and Jacquiline 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, M. 1991. Commentary on Wolford, Taylor, and Beck: The conjunction fallacy? *Memory and Cognition*. 19. 412-414.

Baron, J. 2000. Normative and prescriptive implications of individual differences. [response to K. E. Stanovich & R.F. West] pp. 668-669 in *Behavioral and Brain Sciences*. 23. 645-726.

Becker, G. and Stigler, G. 1977. De gustibus non es disputandum. *American Economic Review*. 67(2). 76-90.

Benbow, C.P. and Stanley, J.C. 1980. Sex differences in mathematical ability: Fact or artifact? *Science*. 210. 1262-1264.

Benjamin, D.J. and Shapiro, J.M. 2005. Does cognitive ability reduce psychological bias? Working paper. Harvard University.

Byrnes, J.P., Miller, D.C., and Schafer, W.D. 1999. Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*. 125(3). 367-383.

Cacioppo, J.T. and R.E. Petty. 1982. The need for cognition. *Journal of Personality and Social Psychology*. 1982. 42(1). 116-131.

Cacioppo, J.T., Petty, R.E., Feinstein, J.A., and W.Blair G. Jarvis. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*. 119(2). 197-253.

Cacioppo, J.T., Petty, R.E., and Kao, C.F. 1986. The efficient assessment of need for cognition. *Journal of Personality Assessment*. 48. 306-307.

Chaiken, S. and Trope, Y. 1999. *Dual-Process Theories in Social Psychology*. New York: Guilford Press.

Donkers, B., Melenberg, B. and Van Soest, A. 2001. Estimating risk attitudes using lotteries: A large sample approach. *The Journal of Risk and Uncertainty*. 22(2). 165-195.

Frederick, S. 2002. Automated choice heuristics. *in* T. Gilovich, D. Griffin, and D. Kahneman [eds]. *Heuristics & Biases: The Psychology of Intuitive Judgment*. New York. Cambridge University Press. pp. 548-558

Frederick, S., Loewenstein, G., and O'Donoghue, T. 2002. Time discounting and time preference: a critical review. *Journal of Economic Literature*. 40. 351-401.

Funder, D.C. and Block, J. 1989. The role of ego-control, ego-resiliency, and IQ in delay of gratification in adolescence. *Journal of Personality and Social Psychology*. 57(6). 1041-1050.

Halpern, D.F. 1986. *Sex Differences in Cognitive Abilities*. Hillsdale, NJ: Erlbaum.

Hedges, L.V. and Nowell, A. 1995. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*. 269. 41-45.

Hilton, D.J. 1995. The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*. 118. 248-271.

Hyde, J., Fennema, E., and Lamon, S. 1990. Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*. 107. 2. 139-155.

Jensen, A.R. 1998. *The g Factor: The Science of Mental Ability*. Praeger. 648 pp.

Kahneman, D. and 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*. Cambridge University Press.

Kahneman, D. and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*. 47. 263-291.

Lubinski, D. and Humphreys, L. 1997. Incorporating general intelligence into epidemiology and the social sciences. *Intelligence*. 24(1). 159-201.

McClure, S.M., Laibson, D.I., Loewenstein, G. and Cohen, J. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science*. 306. 503-507.

Mischel, W. 1974. Processes in delay of gratification. In L. Berkowitz (Ed.). *Advances in Experimental Social Psychology*. Volume 7. pp. 249-292. San Diego, CA: Academic Press.

Monterosso, J., Ehrman, R., Napier, K.L., and O'Brien, C. P., and Childress, A.R. 2001. Three decision-making tasks in cocaine-dependent patients: do they measure the same construct? *Addiction*. 96. 1825-1837.

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

Rabin, M. 2000. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*. 68(5) 1281-1292.

Rae, J. 1834. *The New Principles of Political Economy*. Reprinted in 1905 as *The Sociological Theory of Capital*. New York: Macmillan.

Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley.

Shoda, Y., Mischel, W. & Peake, P.K. 1990. Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*. 26(6). 978-986.

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

Slovic, P. and Tversky, A. 1974. Who accepts Savage's Axiom? *Behavioral Science*. 19(4). 368-373.

Stanovich, K.E. and West, R.F. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*. 23. 645-726.

Sternberg, R.J. 2000. The ability is not general, and neither are the conclusions. [response to K. E. Stanovich & R.F. West] pp. 697-698 in *Behavioral and Brain Sciences*. 23. 645-726.