Journal of Behavioral Decision Making
J. Behav. Dec. Making, **18**: 1–27 (2005)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/bdm.481

Decision-making Competence: External Validation through an Individual-differences Approach

ANDREW M. PARKER[1]* and BARUCH FISCHHOFF[2]
[1]*Virginia Polytechnic Institute and State University, USA*
[2]*Carnegie Mellon University, USA*

ABSTRACT

This study asks to what extent (a) individuals show consistent performance differences across typical behavioral decision-making tasks, and (b) how those differences correlate with plausible real-world correlates of good decision making. Seven tasks, chosen to span the domain of decision-making skills, were administered to participants in an ongoing longitudinal study providing extensive social, psychological, and behavioral measures. Performance scores on individual tasks generally showed small, positive inter-task correlations. An aggregate measure of decision-making competence (DMC) was appropriately correlated with plausible sources, concomitants, and outcomes of good decision making, suggesting the underlying construct's external validity. Higher DMC scores were associated with more intact social environments, more constructive cognitive styles, and fewer 'maladaptive' risk behaviors. In each case, DMC adds to the predictive validity of general measures of cognitive ability. These results suggest that poor decision making on common laboratory tasks is related to real-world antecedents and consequences of poor decision making. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS individual differences; decision making; competence; late adolescence; judgment; external validity

INTRODUCTION

Every day, people face decisions in domains as diverse as choosing among shampoos, stocks, medical treatments, and friends. When people have not learned what to do through trial and error, they need a suite of generally applicable decision-making skills. These include extracting relevant information, applying general values in specific settings, and integrating these pieces with a coherent decision rule.

* Correspondence to: Andrew M. Parker, Department of Marketing (0236), Virginia Tech, Blacksburg, VA 24061, USA.
E-mail: anparker@vt.edu

Contract/grant sponsors: National Institute on Drug Abuse, USA; National Institute of Allergies and Infectious Diseases, USA; contract/grant number: SBR-9521914.
Contract/grant sponsor: National Science Foundation, USA; contract/grant number: SES-0213782.

Behavioral decision research has often found deficiencies in these skills. These limits are reflected in such phenomena as preference reversals, anchoring and (insufficient) adjustment, sunk costs effects, poorly calibrated confidence judgments, and hindsight bias (Dawes & Hastie, 2001; Kahneman et al., 1982; Yates, 1990). This research is primarily experimental. Its tasks have evolved over time, as investigators responded to emerging theoretical issues and concerns about possible confounds in previous studies. As a result, the nuances of many tasks are relatively well understood. However, investigators disagree about the level of external validity, pointing to the unnatural tasks and artificial environments used in the lab (Gigerenzer et al., 2000; Klein, 1999).

In most behavioral decision research, researchers' primary interest has been general cognitive processes. That focus has not only diverted attention from individual differences (Lopes, 1987), but has also led to constant evolution and experimental variation of tasks—preventing the standardization needed for psychometric evaluation. Nonetheless, many decision-making tasks reveal considerable variance across respondents. Stanovich and West (1998) have studied this variance most extensively, finding significant positive correlations between individuals' performance on some tasks (e.g., measures of under/overconfidence, hindsight bias, and "false" consensus effects). Stanovich and West (2000) characterized these correlations as reflecting a "positive manifold" of decision-making performance.

If there is a common factor underlying performance on behavioral decision-making tasks, then one can ask more orderly questions about their external validity. Stanovich and West argued that decision-making competence should be related to cognitive ability and, indeed, found significant correlations between a composite measure of cognitive ability (combining scores on the Scholastic Aptitude Test, Raven Advanced Progressive Matrices, and Nelson–Denny Reading Test) and resistance to two judgment problems: overconfidence and hindsight bias. They also predicted that performance on decision-making tasks would be related to cognitive styles such as open-minded thinking (Baron, 1988). However, neither overconfidence nor hindsight bias was related to the thinking disposition composite (TDC), a measure of cognitive style that was correlated positively with performance on other reasoning tasks (Stanovich & West, 1998). Levin and colleagues found positive correlations between need for cognition (Cacioppo & Petty, 1982), another cognitive style, and two aspects of decision-making performance: resistance to framing effects (Smith & Levin, 1996) and being an "adaptive decision maker" (Levin et al., 2000; Payne et al., 1993). Wolfe and Grosch (1990) reported correlations between overconfidence and individual-difference measures of affect (including optimism), social cognition (need for cognition, self-monitoring, self-efficacy), and cognitive ability (SAT). Finucane et al. (2002) found that young adults performed better than older ones in understanding decision-relevant information and making consistent judgments across contexts—a difference that is possibly related to many differences with age. With each age group, decision-making ability correlated positively with education and a rational-vigilant decision style (as reflected in self-reports). Crawford and Stankov (1996) also found that overconfidence on intelligence tests increased with age. Mixed results are found when correlating overconfidence with personality traits (Stankov & Crawford, 1996). In his presidential address to the Society for Judgment and Decision Making, Levin (1999) argued, "We are now in a position to establish some of the links between stable person characteristics and decision processes."

The research reported here extends this work by administering a battery of seven behavioral decision-research tasks to a group of young adults that is unusually well characterized on measures relevant to those tasks' external validity. The seven decision-making tasks were chosen to represent core decision-making skills. Responses are initially analyzed in terms of the decision-making measures' internal validity, namely consistency and stability over time. Their external validity is then evaluated in terms of other measures of cognitive ability and style that should be related to decision-making competence—relations that have sometimes been found in previous studies, with more restricted sets of decision-making tasks. In addition, the current sample allows correlating decision-making performance with plausible "real-world" antecedents and consequents, including social class, family structure, psychopathology, and self-reported risk behaviors. Such tests of external validity have not been possible in previous research. Initial analyses use the seven behavioral decision-research tasks as separate predictors. The largest factor, extracted from a simple factor analysis, is then used to examine the construct validity of a general measure of decision-making competence (DMC).

MEASURING DECISION-MAKING COMPETENCE: DEFINING THE DOMAIN

Accounts of decision-making processes typically point to four fundamental skills: assessing beliefs; assessing values; combining beliefs and values in order to identify choices; and having a meta-cognitive understanding of one's abilities (Edwards, 1954; Raiffa, 1968). For each skill, performance can be defined in terms of *accuracy*, relative to an external criterion (e.g., Do risk judgments match actuarial estimates?), or internal *consistency* (e.g., Do preferences on one problem contradict those on another?) (Dawes & Hastie, 2001; Yates, 1990). In order to capture an overall picture of decision-making competence, we selected tasks representing each of these skill sets, with a mixture of accuracy and consistency standards. Some of these often-complex tasks may tap skill sets other than the one with which they are primarily identified. As a result, our primary goal was to ensure that each skill set was represented in our set of tasks, rather than to try to isolate it. The next sections discuss each skill's conceptualization and operationalization, placing tasks in what might be thought of as their primary skill set. Although we focus on cognitive skills, the approach could be extended to affective or motivational decision-making skills.

Table 1 presents the seven tasks, in terms of skill sets, criterion (consistency or accuracy), and action (judgment or choice). All four possible combinations of criterion and action are represented (although not for each skill set).

Belief assessment

Conceptualization

Belief assessment involves judging the probabilities of events occurring (or, more generally, of statements being true) (Edwards et al., 1963; Fischhoff & Beyth-Marom, 1983). Beliefs play a central role in both normative and descriptive theories of decision making (Camerer, 1992; Kahneman & Tversky, 1979; Schoemaker, 1982) and risk behavior (Vlek & Stallen, 1981; Yates, 1992). Probability judgments have been found to have intertemporal reliability, a necessary condition for psychometric validity (e.g., Peterson et al., 1965; Wallsten & Budescu, 1983; Whitcomb et al., 1993). Many studies have examined the accuracy and

Table 1. The seven DMC component measures

DMC component measure				
Name	Description	Skill sets[a]	Criterion	Action
Consistency in risk perception	Number of consistent risk judgments	BA	Consistency	Judgment
Recognizing social norms	Rank correlation between judged proportion and true proportion	BA, VA	Accuracy	Judgment
Resistance to sunk costs	Number of times resisted sunk cost	VA	Accuracy	Choice
Resistance to framing	Number of consistent choice pairs	VA, INT	Consistency	Choice
Applying decision rules	Number of correct answers	INT	Accuracy	Choice
Path independence	Number of consistent choice pairs	VA, INT	Consistency	Choice
|Under/overconfidence|	1 − |Difference between mean confidence judgment and percent correct|[b]	BA, MC	Accuracy	Judgment

[a]BA = belief assessment; VA = value Assessment; INT = integration; MC = metacognition.
[b]The measure described in the text was transformed in this way in order to make higher scores reflect better performance.

consistency of risk beliefs (e.g., Fischhoff et al., 2002; Fischhoff & MacGregor, 1983; Lichtenstein et al., 1978; Slovic, 2001).

Operationalization

We assessed this skill with two tasks. The first elicits probability judgments for 20 events chosen to have roughly similar familiarity for diverse individuals. They are adapted from the expectations module of the 1997 National Longitudinal Study of Youth (1997; Fischhoff et al., 2000).[1] Each item asks respondents to judge the probability of an event happening to them in a specified time period. The set begins with ''eating pizza'' and ''getting the flu,'' then continues with ''getting [someone] pregnant,'' ''being the victim of a violent crime,'' ''using illegal drugs,'' and ''dying,'' among other events. Responses are marked on a linear, numerical scale, anchored at 0% = ''no chance'' and 100% = ''certainty.'' Ten events were in formally related pairs, including proper subsets/supersets, conjunctions and disjunctions, and conditional probabilities (e.g., ''percent chance that you will die in the next year' is a proper subset of ''percent chance that you will die between now and when you turn 30''—hence should be no larger). *Consistency in risk perception* scores equally the number of logically consistent pairs of judgments (from 0 to 5). Although we lacked the personal information needed to evaluate individuals' beliefs for accuracy, that is, in principle, possible (Fischhoff et al., 2000).

The second task assesses individuals' ability to judge likelihood, using a strategy following Jacobs et al. (1995) and Loeber (1989). Respondents first answered 16 questions, asking whether they ''think it is sometimes OK'' to engage in various negative behaviors (e.g., ''using your fists to resolve a conflict''). They then estimated how many ''out of 100 people your age'' endorse each position. *Recognizing social norms* scores equal the within-respondent rank-order correlation (from -1 to $+1$) between estimated and actual social norms (defined by pooling responses across the sample). Several unrelated tasks separated the two sets of questions, in order to make the judgments as independent as possible (within the constraints of a single session).

Value assessment

Conceptualization

Well-articulated values should be sensitive to relevant task changes and insensitive to irrelevant ones (Fischhoff, 1991). Some of the most troubling behavioral decision-research deviations from rationality have involved inconsistent preferences for formally equivalent choices (Kahneman & Tversky, 1979). Avoiding such inconsistency in this respect will be our main measure of value-assessment performance.

In decision theory, values are a matter of individual taste. Their accuracy cannot be evaluated in terms of an external standard. Nonetheless, understanding social norms is necessary for having values that reflect (or reject) what others think. Perceived social norms are central to the theory of reasoned action (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) and play a key role in both ''false'' consensus effects (Dawes, 1990; Hoch, 1987) and pluralistic ignorance (Miller & McFarland, 1987; Prentice & Miller, 1993). As a result, recognizing social norms will be a secondary measure of performance in value-assessment.

Operationalization

Insensitivity to irrelevant task features was assessed in two ways. One involved comparing responses to formally equivalent (or ''extensional'') forms of simple choices (Tversky & Kahneman, 1981; Fischhoff, 1983; Levin et al., 1998). Frisch (1993) demonstrated that framing effects can be observed within a single session. We used five such pairs of questions, with members separated by approximately half an hour of unrelated tasks. The first pair, modified from Shafir (1993), asks respondents, at the different times, to *choose* one of

[1]Full instructions and instruments are available from the authors on request.

two alternatives or to *reject* one of the two. Consistency means choosing one option and rejecting the other. The second pair of choices, modified from Linville, Fischer, and Fischhoff (1993), asks respondents to decide whether a condom is acceptable, with efficacy described in terms of either (a) its *success* rate or (b) its (complementary) *failure* rate. Consistency means choosing the same option both times. The third pair, adapted from Roelofsma and Keren (1995), presents choices (a) between receiving $100 tomorrow or $120 in 4 weeks and (b) between receiving $100 in 26 weeks or $120 in 30 weeks. Consistency means accepting or rejecting the $20 compensation both times (i.e., whether the four-week delay begins in 4 weeks or 30 weeks). The fourth pair, from Tversky and Kahneman (1988), presents equivalent medical treatments in terms of either (a) lives saved or (b) lives lost, a labeling change that should make no difference. The last pair, from Fischhoff (1993), requires choosing between a gamble with negative consequences and a certain expenditure of equal expected value, described as either (a) a sure loss or (b) an insurance payment—a formally irrelevant change. *Resistance to framing* equals the number of consistent choices across these problem pairs (0–5).

As a second measure of insensitivity to irrelevant task features, we considered commitment to sunk costs (Arkes & Blumer, 1985). Normatively, prior investments should be ignored, so that decisions reflect just future consequences. Respondents received two problems, each offering the choice between continuing an action in which an investment had been made and switching to one with better consequences. The problems are adapted from ones in Baron et al. (1993) and Dawes and Hastie (2001). *Resistance to sunk costs* equals how often respondents reject the sunk-cost option (0–2).

As mentioned, understanding peers' values might be considered as a form of accuracy in value assessment, captured here by recognizing social norms.

Integration

Conceptualization

Integration involves combining beliefs and values coherently when making decisions. Better integration processes should result in selecting more appropriate decision rules, then executing them more accurately and consistently, in the face of irrelevant shifts in decision structure (analogous to insensitivity to irrelevant value considerations).

Operationalization

We assessed integration ability in one way that addresses accuracy and another that addresses consistency. Our accuracy measure assessed respondents' ability to apply a specified decision rule. It used seven questions adapted from Payne, Bettman, and Johnson's (1993) studies of the relative difficulty of implementing various decision rules. Each asks respondents to apply a specified rule in a multi-attribute choice. Figure 1 displays one of five choices involving Walkmen. The specified (lexicographic) rule should lead to option C. Two other questions paralleled Walkmen questions in structure; but used cover stories that might be anxiety-provoking, hence reduce cue utilization and encourage simpler decision strategies (Easterbrook, 1959; Janis & Mann, 1976; Luce, 1998; Mano, 1992). One cover story involved sex, the other drug use. *Applying decision rules* equals the number of correct applications (0–7). Three initial questions assessed respondents' ability to read the tables (e.g., ''Which Walkman is best in comfort of headphones?'').[2]

The rational-choice axiom of *path independence* requires indifference to the sequence of events leading to an outcome. We assessed compliance by comparing responses to 12 pairs of choices, modeled on stimuli common in experimental studies. Each poses a hypothetical choice between either (a) a gamble and a sure

[2]Five respondents got one preliminary question wrong (no one got more than one wrong). Of these, three answered all seven of the applying-decision-rules questions correctly, one got four wrong, and one got five wrong. Excluding the latter two (on the grounds that they may have misunderstand the tables) had little effect on the analyses.

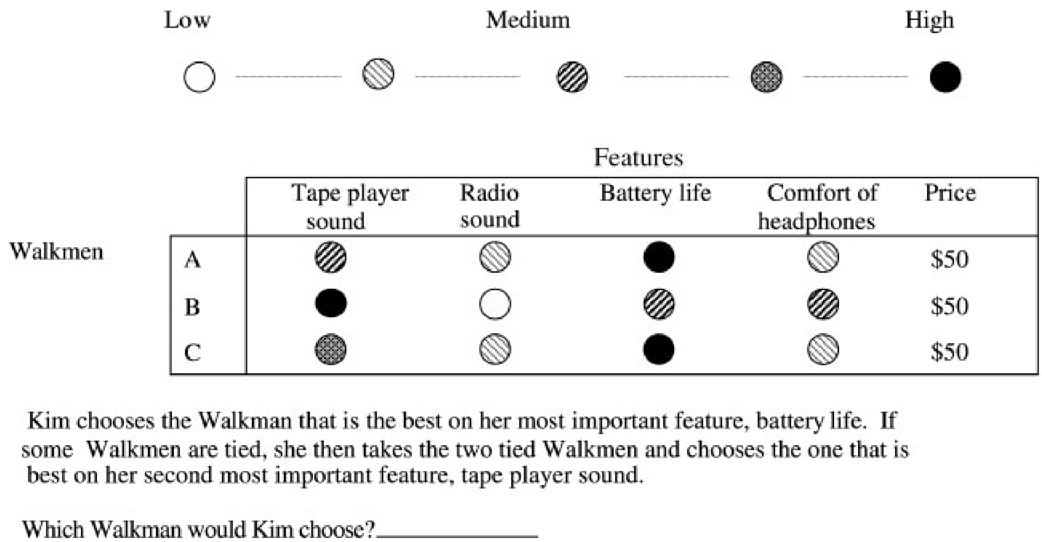


Figure 1. sample question from the applying decision rules task.

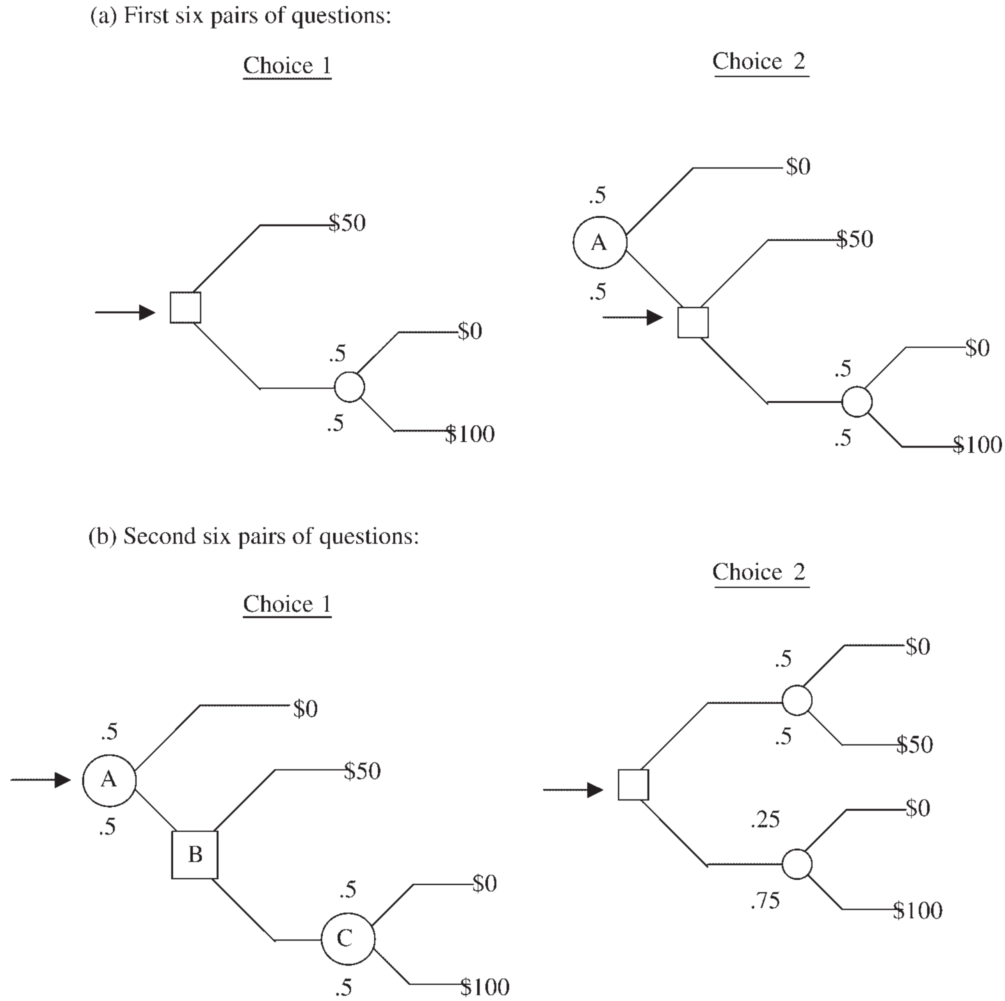
thing or (b) two gambles, leading to equivalent outcomes. Half had the structure shown in Figure 2a. Choice 1 is between a sure gain of $50 and a coin toss with equal expected value. Choice 2 offers the same options, but says that heads came up on the preceding flip (A) of the same coin. Assuming independent flips, respondents should choose the same option each time. Figure 2b shows the structure of the other questions. Choice 1 is made prior to a coin flip (A) that pays $0, if heads, or a choice (B), if tails. Choice B is between a sure $50 and a coin flip (C) of equal value. The formally equivalent Choice 2 is between a single-flip gamble (paying $0 or $50) and a double-flip gamble (paying $100 if two heads, otherwise $0) (Plous, 1993). Half of the questions in each set involve just losses, half just gains. *Path independence* equals the number of consistent paired choices (0–12). Researchers have typically found little cross-situational consistency in risk-seeking behavior (Bromiley & Curley, 1992; Slovic, 1962)—although Weber, Blais, and Betz (2002) recently found that such consistency improved after controlling for risk judgments. The probabilities in our tasks are determined by a familiar event (coin flips), which should encourage consistent risk-taking preferences. We know of no studies addressing individual differences in following choice axioms.

Although we considered resistance to framing primarily as a measure of value assessment, those tasks also require integrating beliefs and values. Conversely, path independence may be conceived as consistency in evaluating complex gambles.

Metacognition

Conceptualization

Knowing the extent of one's competence is critical to effective decision making. People with unwarranted confidence may undertake tasks beyond their abilities, underutilize available assistance, and neglect signs that decisions are going awry. People with insufficient confidence may needlessly hesitate, defer to others, or doubt their ability to identify sound courses of action. Metacognitive processes have been studied in terms of confidence assessment (sometimes measured as "calibration"), process management, and cognitive control.



Note: The arrows (⟶) indicate the location of the decision maker. Squares indicate choices; circles are coin flips.

Figure 2. Comparisons in the path independence task.

Operationalization

In a typical confidence-assessment problem, respondents (a) decide whether a statement is true and (b) express their confidence (between 50% and 100%) that this choice is correct. Many studies have found modest correlations between confidence and knowledge (Keren, 1991; Lichtenstein & Fischhoff, 1977; Yates, 1990). Confidence judgments show good reliability (Murphy, 1997; Wallsten & Budescu, 1983; Whitcomb et al., 1993). Confidence judgments show stable individual differences across tasks, despite variation in accuracy (Pallier et al., 2002; Stankov, 1998; Stankov & Crawford, 1997; West & Stanovich, 1997; Wolfe & Grosch, 1990). In a study with adolescents (Parker et al., 2000), both confidence and accuracy were internally consistent across items (Cronbach alphas of 0.94 and 0.76, respectively).

Respondents received 42 questions, each asking whether a statement was true or false and the probability of that choice being correct. One third each considered (a) general knowledge (e.g., ''robins' eggs are orange''), (b) sex and AIDS (e.g., ''you can only get the AIDS virus (HIV) from someone who is gay''), and (c) drugs or alcohol (e.g., ''alcohol kills brain cells''). *Under/overconfidence* equals 1 minus the absolute value of the difference between mean confidence judgment and proportion of correct choices.[3]

HYPOTHESES

Internal validity: Relationships among DMC component measures

Hypothesis 1

Each DMC component measure is internally consistent.

Past research on each task has involved many variations, seeking to examine different aspects of a common process. If that assumption is correct, then performance on one item, within each task, should correlate positively with performance on other items.

Hypothesis 2

A single factor captures much of the variance in the seven DMC component measures.

In principle, these decision-making skills could be independent of one another. However, the clusters identified in previous studies of individual differences overlap one another. Moreover, the skills reflect a common conception of decision making, hence should, arguably, support one another. For example, making consistent choices should be easier for those who can apply decision rules; metacognition should facilitate the learning processes associated with acquiring other skills. Hence, we expect the seven DMC component measures to share a common factor. That does not preclude these tasks also measuring distinct competencies, leading to the next hypothesis.

Hypothesis 3

DMC factors should correspond to the constructs guiding the task selection.

If there are distinct decision-making competencies, then they should correspond to one of the three theoretical distinctions made above: (a) skill set (belief assessment, value assessment, integration, or metacognition); (b) criterion (accuracy or consistency); and (c) action (judgment or choice). If both Hypotheses 2 and 3 are supported, then the relationship between the common factor and the subfactors becomes an empirical question. The common factor could be related to each of the distinct factors or it could be strongly related to one subfactor, arguably representing the key aspect of decision-making competence. As will be seen, our data set, although unique in many respects, has too small a sample of individuals for more than suggestive exploration of these latter questions.

Nomological validity: DMC measures and external measures

All measures in this section are drawn from an ongoing longitudinal study, conducted at the Center for Education and Drug Abuse Research (CEDAR), described more fully below. Respondents were 18–19 years old when they completed the DMC tasks. The CEDAR tasks were done at the ages shown in Table 5. The

[3]Absolute under/overconfidence was chosen from the many possible performance measures because of its intuitive appeal and modest metric assumptions (Yates, 1990). In addition, it does not condition on confidence (unlike the calibration and discrimination indices) and is monotonic in performance (unlike raw under/overconfidence).

temporal structure of these measures reflects CEDAR's design, rather than our concerns, some of which would have benefited from more contemporaneous assessment. Nonetheless, the richness of the CEDAR database still provides a unique opportunity to assess the external validity of traditional laboratory decision-making tasks, using years of in-depth data.

Hypothesis 4

DMC correlates positively with cognitive ability.

The seven behavioral decision-making tasks focus on cognitive skills. As a result, DMC should be related to other measures of cognitive ability. After analyzing many standardized measures of cognitive abilities, McGrew (1997) distinguished between ''crystallized'' cognitive abilities (*Gc*), or *knowledge*, and ''fluid'' cognitive abilities (*Gf*), or *reasoning*.

As a measure of knowledge, we used the vocabulary portion of the WISC-R (Wechsler, 1972), which loads strongly on *Gc*, across many cognitive domains. It is also relatively independent of other aspects of executive cognitive function, unlike other aspects of the Wechsler verbal battery (Giancola et al., 1998).

Gf is typically assessed with non-verbal tasks, such as the block-design, picture-arrangement, and object-assembly tests from the Wechsler intelligence scales (WISC-R; Wechsler, 1972; WAIS-R; Wechsler, 1981), or the Porteus maze, vigilance, motor restraint, and Stroop tasks. As a measure of *Gf*, we used the executive cognitive functioning (ECF) test developed by Giancola et al. (1996), as part of the CEDAR project.[4] It reflects Giancola et al.'s (1998) conceptualization of ECF as ''a 'higher order' cognitive construct involved in planning, initiating, and regulating goal-directed behavior.'' It tests such abilities as attention, strategic-goal planning, hypothesis generation, temporal response sequencing, spatial ability, and working memory. Although there is some conceptual overlap between ECF and DMC, ECF is conceptualized as being more abstract and less context-bound (Giancola, personal communication).[5]

Hypothesis 5

DMC correlates positively with measures of constructive and introspective cognitive styles.

As mentioned, Baron (1988) and Stanovich and West (1998) argue that DMC should involve constructive, introspective, and complex styles of thinking. It should also be related to self-consciousness (concern for how others view one) and self-monitoring (critiquing one's behavior in light of social and situational factors). We measured these *cognitive styles* with the polarized-thinking portion of the constructive thinking inventory (Epstein & Meier, 1989; Katz & Epstein, 1991), the self-consciousness scale (Fenigstein et al., 1975), and the self-monitoring scale (Graziano et al., 1987; Snyder, 1974; Snyder & Cantor, 1980).

We also used the behavioral coping component of the constructive thinking inventory. It measures whether individuals endorse such problem-solving skills as planning and separating complex problems into smaller ones. Thus, it assesses attitudes towards skills that several of the DMC measures address behaviorally (recognizing that individuals may not practice, or even possess, skills that they endorse).

---

[4]ECF uses a factor-score composite of (a) the Porteus Maze Test (Porteus, 1965), which requires navigating eight mazes without lifting the pencil from the paper; (b) the vigilance task (Pelham et al., 1992), which requires pressing the space bar when target letters appear in a sequence of matrices of letters and boxes on a computer screen; (c) the Motor Restraint Task (Parsons et al., 1972), which requires tracing a 180° arc as slowly as possible; (d) the forbidden toy task (Cole et al., 1993; Silverman & Ragusa, 1992), which requires filling out worksheets in a room filled with interesting toys that the participant is not allowed to play with; and (e) the block design test from the Wechsler Intelligence Scale for Children—Revised (Wechsler, 1972), which involves organizing blocks into designs matching pictoral models.

[5]Baron (1985) draws a similar distinction between relatively unchangeable *capacities*, including ECF-like abilities, and more malleable *dispositions*, such as thinking styles. Stanovich and West (1998) treat capacities and dispositions as different levels of analysis. Overton (1990) distinguishes between *competence* (more similar to ECF than DMC) and *activation-utilization*.

Hypothesis 6

DMC correlates negatively with "maladaptive" risk behavior.

Decisions to engage in *risk behaviors* (e.g., drug use, under-age drinking) need not be irrational—if they follow from an individual's beliefs and values in an orderly, consistent fashion. Nonetheless, they would be poor choices in a society that proscribes them and for individuals who fully understand their consequences. As a result, DMC should be associated with avoiding such "maladaptive" risk behaviors (Dryfoos, 1990; Jessor & Jessor, 1977; Resnick et al., 1997).

We used the following CEDAR measures: (a) antisocial disorders, specifically conduct disorder and oppositional-defiant disorder, using a diagnostic instrument derived from Endicott and Spitzer (1978), Spitzer et al. (1987); (b) externalizing behavior, reflecting aggression and directing anger outward, as reported by each respondent's mother at age 14 on the child behavior checklist (Achenback & Edelbrock, 1983); (c) delinquency, as reported by respondents' mothers on the Child Behavior Checklist (Achenback & Edelbrock, 1983); (d) alcohol and marijuana consumption, equal to the self-reported numbers of full drinks and episodes of marijuana use, corrected for age, through the age-16 assessment (Skinner, 1982); and (e) respondents' self-reported number of times having sex and number of sexual partners, in the 12 months before their age-16 assessment.

Hypothesis 7

DMC correlates positively with measures reflecting positive social and family environments.

As a set of learned skills, DMC should be fostered by a positive *developmental environment*, defined as one providing pertinent feedback and modeling (Parrill-Burnstein, 1978; Baron & Hershey, 1988; Jones et al., 1997). Conversely, a critical or inconsistent home environment may undermine young people's confidence in their decision-making abilities (Institute of Medicine, 1999). We assessed these environments with CEDAR's measures of parental substance abuse (Freund, 1990), socioeconomic status (SES; Hollingshead, 1975), social support (Cohen & Hoberman, 1983), and negative peer environment (Tarter, 1991).

CEDAR

Data were collected at the Center for Education and Drug Abuse Research (CEDAR), funded by the National Institutes of Health. The CEDAR sample has approximately 700 Pittsburgh-area youth, participating in a longitudinal study on the etiology of substance use, along with many parents, siblings, and peers. After a two-day initial assessment at 10–12 years old, participants have additional visits every 2–3 years. Each visit repeats part of the baseline protocol, which is repeated almost in its entirety at age 16. The assessments include physical (e.g., health history), psychiatric (e.g., DSM diagnostic history), psychological (e.g., self-monitoring), behavioral (e.g., alcohol and marijuana use), social (e.g., peer environment), and family (e.g., cohesion) functioning. Before the first assessment, respondents are defined as high average risk (HAR), if the father has a history of substance abuse, or low average risk (LAR), if not. HAR youth are over-sampled in order to obtain subsamples of equal size.

The present data are from 110 participants in the fourth CEDAR assessment, at 18–19 years of age. All are male. CEDAR added female subjects after it started; none had reached age 18–19 at the time of our study. The main CEDAR protocol takes about half a day. At its end, 61 individuals were asked to complete the DMC protocol; all agreed. Questionnaires were mailed to 101 individuals who had already completed the fourth assessment; 49 were returned. All tasks are self-administered, using paper and pencil. In order to test reliability, 30 respondents were asked to repeat the DMC protocol 1–6 months after the initial administration; 60% did so.

Table 2. Descriptive performance statistics

DMC component measure*	*N*	Number of items	Potential range	Observed range	Median	Mean	*SD*	Cronbach alpha
Consistency in risk perception	110	5	0, 5	1, 5	5	4.34	0.93	0.50
Recognizing social norms	110	16	−1, 1	−0.15, 0.91	0.58	0.57	0.18	
Yes/no								0.79
percent								0.88
Resistance to sunk costs	109	2	0, 2	0, 2	1	0.80	0.70	0.03
Resistance to framing	108	5	0, 5	1, 5	4	3.68	1.09	0.30
Applying decision rules	110	7	0, 7	2, 7	7	6.23	1.28	0.68
Path independence	109	12	0, 12	0, 12	9	8.12	2.67	0.70
|Under/overconfidence|	110	42	0, 1	0.71, 1.00	0.95	0.94	0.06	0.79

*All DMC component measures are coded such that higher numbers indicate better performance.

RESULTS

Concurrent validity: Relationships among DMC component measures

Hypothesis 1

Each DMC component measure is internally consistent.

Table 2 displays descriptive statistics for the seven DMC component measures, coded so that higher scores represent better performance. On the whole, scores cover their potential ranges, increasing the chances of detecting differences among respondents. Cronbach alpha measures test items' degree of interrelation, potentially reflecting one or more common factors (Cronbach, 1951; Cortina, 1993). It equals the mean of all split-half correlations among items. Low Cronbach alphas mean either unreliable items or ones tapping different processes. Either possibility limits the chances of getting meaningful, stable results from composite measures. For all tasks except recognizing social norms and |under/overconfidence|, Cronbach alpha measures consistency in respondents' tendency to commit errors. For recognizing social norms, alpha is calculated separately for how consistently respondents endorse the proposed behavior (yes/no) and how consistently they offer high estimates of the frequency of others' endorsements (percent). For |under/overconfidence|, internal consistency was computed for the absolute difference between answering each item correctly (scored $[0, 1]$) and the corresponding confidence judgment $[0.5, 1]$. These alpha levels are good for the judgment measures (consistency in risk perception, recognizing social norms, and |under/overconfidence|). The alphas are appreciably poorer for some of the choice tasks (notably, resistance to framing and resistance to sunk costs).[6]

Hypothesis 2

A single factor captures much of the variance in the seven DMC component measures.

Table 3 displays bivariate correlations among the seven DMC component measures. They are moderate to low, with all either positive or slightly negative (mean, median = 0.12). The strongest correlations involve resistance to framing, applying decision rules, and under/overconfidence. Thus, these seven measures reveal a weak positive manifold.

[6] One potential reason for the better alphas among the judgment measures is the larger range of options, allowing for more discrimination among responses. Another possibility is that the judgment measures more uniformly reflect underlying psychological constructs (e.g., beliefs or values), while the choice items rely on behavior dependent on varying situational factors (e.g., cover stories).

Table 3. Pearson correlations between DMC component measures

	Risk perception	Social norms	Sunk costs	Framing	Decision rules	Path independence
Consistency in risk perception	1					
Recognizing social norms	0.18*	1				
Resistance to sunk costs	0.15	0.23**	1			
Resistance to framing	0.22*	0.16	0.02	1		
Applying decision rules	0.12	0.08	0.03	0.24**	1	
Path independence	−0.05	0.03	0.21*	0.05	0.06	1
|Under/overconfidence|	−0.01	0.05	−0.05	0.31***	0.35***	0.17*

*One-sided $p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
Note: mean correlation = 0.12, median = 0.12.

We performed an exploratory factor analysis on *z*-scores of the seven measures, using the principal-components method, because of its appropriateness for exploratory research and modest distributional assumptions (Mardia et al., 1979). Table 4's first column shows loadings for the one-factor model, which accounts for 25.1% of the variance. Thus, a central cross-task measure of DMC captures a moderate portion of performance variance.

Hypothesis 3

DMC factors should correspond to the constructs guiding the task selection.

Table 4's next three columns show the three factors with eigenvalues greater than 1, using oblimin rotation (as there was no a priori reason to believe that these factors are orthogonal). They account for 60.2% of the variance. The largest loadings for each measure are in bold. The first factor reflects primarily resistance to framing, applying decision rules, and under/overconfidence. Their respective task features are (value assessment and integration, consistency, choice), (integration, accuracy, choice), and (belief assessment and metacognition, accuracy, judgment). Thus, this factor reflects some crosscutting aspect of decision-making competence, rather than one of the three distinctions guiding task selection. The largest loadings on the second factor are for recognizing social norms, consistency in risk perception, and resistance to sunk cost, also cutting across the three distinctions. The third factor most strongly reflects path independence and resistance to sunk cost. Thus, the structure revealed by the factor analyses does not cleanly correspond to the distinctions guiding the task selection.

Table 4. Loadings for the one- and three-factor DMC models

	One-factor model	Three-factor model (oblimin rotation)		
Resistance to framing	0.67	**0.67**	0.35	−0.11
Applying decision rules	0.63	**0.71**	0.11	0.01
|Under/overconfidence|	0.62	**0.78**	−0.12	0.19
Recognizing social norms	0.44	0.11	**0.67**	0.16
Consistency in risk perception	0.29	0.15	**0.72**	−0.26
Resistance to sunk costs	0.39	−0.14	**0.55**	**0.60**
Path independence	0.31	0.16	−0.05	**0.84**
Eigenvalue	1.76	1.76	1.31	1.141
Variance explained	25.1%	25.1%	18.8%	16.3%
			(Total = 60.2%)	

Note: **Bold type** indicates factor with largest loading for each variable. It is used twice for resistance to sunk costs, because of the similarity of the loadings on the second and third factors.

Given that the second and third factors are neither strong nor conceptually distinct, we focus subsequent analyses on the forced, one-factor model, closely reflecting the first factor from the unrestricted factor analysis. We use Anderson–Rubin factor scores extracted from this model as a latent-construct measure of decision-making competence (DMC). (These scores were highly correlated ($r = 0.96$) with those from a maximum-likelihood exploratory factor analysis forcing a single-factor solution.) Using Nunnally and Bernstein's (1994) computation for linear combinations of measures, the reliability of DMC is approximately 0.76. The test–retest reliability correlation for the DMC factor score was 0.54 ($p = 0.02$), indicating reasonable stability over time (mean = 3 mos.; range = 1–6 mos.), for the 19 participants who completed it twice.

Nomological validity: DMC and CEDAR antecedents

Analysis strategy

Table 5 presents zero-order Pearson correlations between the individual DMC and CEDAR measures. Given the large number of correlations, an overall p-value was computed for each DMC component within each section of the table (e.g., risk perception with the entire cognitive ability section) using Strube's (1985) method for combining significance levels from non-independent hypothesis tests. Table 6's first column shows parallel analyses for the DMC factor scores.

Table 6's remaining columns distinguish DMC's role from those of vocabulary and ECF. We used two approaches in these analyses. One uses semipartial correlations. Also called part correlations, they parallel the reduction in a regression beta coefficient, after adding a mediating variable to a model.[7] A limit to this procedure is that it assumes no error in measuring the mediating construct. Clearly, neither vocabulary nor ECF satisfies this assumption. The second approach, based in structural equation modeling, tests the *one-mediator null hypothesis* (Birnbaum, 1985). Rejecting this hypothesis indicates an independent link between DMC and the dependent measure. The second or third column of Table 6 has a superscript when this hypothesis is rejected, with *b* representing stronger evidence of a link than *a*. The Appendix provides details. (Asking these questions about the component tests or additional factors seems best postponed to a larger data set.)

Hypothesis 4

DMC correlates positively with cognitive ability.

The first two rows of Table 5 show strong correlations between five of the seven DMC component measures and respondents' scores on the WISC-R vocabulary test and on Giancola et al.'s (1996) measure of ECF. Consistency in risk perception and resistance to sunk cost show little relationship to either of these general cognitive abilities.[8,9]

Table 6's first two rows show the same strong correlations for the overall DMC measure. The semipartial correlations show reduced, but still significant, relationships. There were no statistical differences between

[7]Another possible analysis looks at partial correlations, which remove the variance shared by the partialled variable and each of the other two variables, rather than the semipartial correlations, which only remove variance shared with the other predictor (in this case DMC). For all relationships reported here, the partial correlations were virtually identical to the semipartials.

[8]While the ECF measure, as a weighted average of five tasks, provides a more diverse measure of fluid cognitive abilities than would, say, the WISC-R block-design task alone, it is also less directly comparable to the single WISC-R vocabulary score. Correlations between the seven DMC component measures and the WISC-R block-design score alone are 0.06, 0.20, 0.19, 0.21, 0.38, 0.31, and 0.31, respectively. Overall, these correlations are somewhat stronger for recognizing social norms, resistance to sunk cost, and path independence, while weaker for the three tasks that showed the strongest correlations with the CEDAR measures and loaded most strongly on the overall DMC factor: Resistance to framing, applying decision rules, and under/overconfidence. Hence, we will focus on ECF rather than the single block-design score, since controlling for ECF would seem to provide a more conservative test of the nomological validity of these three DMC components and the overall DMC measure.

[9]Semipartial correlations between the DMC component scores and vocabulary, controlling for ECF were 0.14, 0.40, 0.08, 0.19, 0.20, 0.00, and −0.01, respectively. Controlling for vocabulary, the correlations with ECF were −0.02, −0.12, 0.03, 0.11, 0.27, 0.25, and 0.33, respectively.

Table 5. Correlations between DMC component measures and CEDAR variables

DMC component[d] correlated with	Risk perception	Social norms	Sunk cost	Framing	Decision rules	Path independence	|Under/overconfidence|
Cognitive ability							
Vocabulary[a]	0.14	0.40	0.12	0.29	0.40	0.13	0.19
ECF[a]	0.06	0.10	0.08	0.24	0.43	0.25	0.38
*Overall**	$p = 0.12$	$p = 0.001$	$p = 0.13$	$p = 0.001$	$p < 0.0001$	$p = 0.01$	$p = 0.0002$
Cognitive style							
Polarized thinking[c]	−0.09	−0.12	−0.06	−0.24	−0.18	−0.24	−0.20
Self-consciousness[b]	0.18	0.12	0.11	0.13	0.09	0.09	0.01
Self-monitoring[c]	0.10	0.12	−0.01	0.23	0.10	−0.02	0.17
Behavioral coping[c]	0.06	−0.01	−0.10	0.27	0.25	0.13	0.31
Overall	$p = 0.03$	$p = 0.06$	$p = 0.40$	$p = 0.0001$	$p = 0.003$	$p = 0.03$	$p = 0.001$
Risk behavior							
Antisocial disorders[c]	−0.20	0.17	0.04	−0.06	−0.14	−0.13	−0.27
Externalizing behavior[b]	−0.15	−0.02	−0.01	−0.18	−0.27	−0.01	−0.27
Delinquency[b]	0.02	−0.02	−0.08	−0.07	−0.32	−0.10	−0.30
l*n*(lifetime # of drinks)[c]	−0.28	0.06	−0.01	−0.13	−0.04	0.01	−0.22
l*n*(lifetime # of drugs)[c]	−0.14	0.14	−0.03	−0.21	−0.15	−0.02	−0.30
l*n*(# times had sex)[c]	−0.16	0.05	0.00	0.01	−0.22	−0.13	−0.37
l*n*(# sexual partners)[c]	−0.26	−0.03	.03	−0.06	−0.26	−0.10	−0.34
Overall	$p = 0.01$	$p = 0.76$	$p = 0.45$	$p = 0.09$	$p = 0.005$	$p = 0.19$	$p < 0.0001$
Social and family influences							
Risk status (HAR = 1; LAR = 0)[a]	−0.19	−0.14	−0.13	−0.20	−0.21	0.01	−0.23
SES[a]	0.21	0.24	0.03	0.14	0.31	0.03	0.22
Social support[c]	−0.08	−0.09	−0.16	−0.11	−0.25	−0.12	−0.22
Positive peer environment[c]	0.16	0.04	−0.01	0.22	0.21	−0.05	0.38
Overall	$p = 0.01$	$p = 0.06$	$p = 0.49$	$p = 0.02$	$p = 0.01$	$p = 0.72$	$p = 0.002$

*Overall p-values were computed using Strube's (1985) method for combining significance levels from non-independent hypothesis tests. All reported ps are one-sided. The correlation between CEDAR measures was used as an estimate of the correlation between test statistics. A conservative Bonferroni correction on the 57 tests presented here and in Tables 6 and 7 would convert an individual $\alpha = 0.05$ into an $\alpha = 0.0009$. If one were interested in an individual correlation, approximate cutoffs would be $r = 0.16$, $p < 0.05$; $r = 0.22$, $p < 0.01$; $r = 0.29$, $p < 0.001$.

[a]Assessed at baseline, age 10–12; [b]Assessed at approximately age 14; [c]Assessed at approximately age 16; [d]Assessed at age 18–19.

Table 6. Correlations between DMC and CEDAR variables

DMC correlated with	Pearson r	Semipartial correlation, controlling for: Vocabulary	ECF	Vocabulary & ECF
Cognitive ability				
Vocabulary	0.50	—	0.28	—
ECF	0.48	0.26	—	—
*Overall**	$p < 0.0001$	$p = 0.0009$	$p = 0.0008$	—
Cognitive style				
Polarized thinking	−0.34	−0.20	−0.24	−0.19
Self-consciousness	0.20	0.14[b]	0.05	0.11
Self-monitoring	0.24	0.29[b]	0.30[b]	0.32
Behavioral coping	0.32	0.27[a]	0.28[a]	0.26
Overall	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
Risk behavior				
Antisocial disorders	−0.19	−0.18[b]	−0.05	−0.09
Externalizing behavior	−0.32	−0.28[b]	−0.18	−0.20
Delinquency	−0.29	−0.28[b]	−0.18	−0.21
l*n*(lifetime # of drinks)	−0.18	−0.22[b]	−0.15	−0.18
l*n*(lifetime marijuana use)	−0.25	−0.30[b]	−0.20	−0.25
l*n*(# times had sex)	−0.24	−0.30[b]	−0.21	−0.27
l*n*(# sexual partners)	−0.30	−0.33[b]	−0.29[a]	−0.31
Overall	$p = 0.0004$	$p = 0.0002$	$p = 0.009$	$p = 0.002$
Social and family influences				
Risk status (HAR = 1; LAR = 0)	−0.35	−0.27	−0.23	−0.21
SES	0.35	0.20	0.21	0.15
Social support	−0.30	−0.21	−0.23	−0.19
Positive peer environment	0.33	0.35[b]	0.32[a]	0.35
Overall	$p = 0.0002$	$p = 0.002$	$p = 0.006$	$p = 0.007$

*Overall p-values were computed using Strube's (1985) method for combining significance levels from nonindependent hypothesis tests. All reported ps are one-sided. The correlation between CEDAR measures was used as an estimate of the correlation between test statistics. A conservative Bonferroni correction on the 57 tests presented here and in Tables 5 and 7 would convert an individual $\alpha = 0.05$ into an $\alpha = 0.0009$. If one were interested in an individual zero-order correlation, approximate cutoffs would be $r = 0.16$, $p < 0.05$; $r = 0.22$, $p < 0.01$; $r = 0.29$, $p < 0.001$. For a semipartial correlation, approximate cutoffs would be $r = 0.18$, $p < 0.05$; $r = 0.25$, $p < 0.01$; $r = 0.32$, $p < 0.001$.
[a]Test A rejects the one-mediator null hypothesis.
[b]Test B rejects the one-mediator null hypothesis.

either the two zero-order correlations (between DMC and the other measures) or the two semipartial correlations (the vocabulary–ECF correlation is 0.51).

Hypothesis 5

DMC correlates positively with measures of constructive and introspective cognitive styles.

As seen in the cognitive style section of Table 5, the DMC component measures have the anticipated negative correlation with polarized thinking and, generally, the anticipated positive correlations with the other measures. The strongest overall relationships with these cognitive style variables involved resistance to framing, applying decision rules, and |under/overconfidence|.

As seen in Table 6, DMC had the same, anticipated correlations. The semipartial correlations suggest that both measures of cognitive ability somewhat mediate some of these relationships, but overall the correlations remain strong. Tests of the one-mediator null hypothesis also indicate direct links between DMC and cognitive-style variables. Thus, DMC has significant relationships with cognitive style, independent of its correlations with these two measures of cognitive ability.

Hypothesis 6
DMC correlates negatively with reported "maladaptive" risk behavior.

Table 5's third section shows generally negative relationships between performance on the DMC component tasks and the measures of maladaptive risk behaviors, despite those having been assessed 2–5 years previously. (Natural logarithms were used with the final four behavioral measures, in order to reduce positive skew.) The strongest overall relationships are with applying decision rules and |under/overconfidence|. Recognizing social norms, path independence, and resistance to sunk cost are, however, largely unrelated to risk behavior.

The overall DMC scores also correlate negatively with each measure of risk behavior (Table 6). The semipartial correlations reveal that most of these correlations remain, after considering vocabulary and ECF, both separately and together. Vocabulary tends to suppress slightly the relationship between DMC and risk behavior, while ECF somewhat mediates it. Tests of the one-mediator null hypothesis show a similar picture: including vocabulary in the model leaves strong links between DMC and the risk behaviors; including ECF reduces them. For the four lifetime behavior counts, the one-mediator null hypothesis is almost rejected with the more conservative Test B (see Appendix). That is, controlling for DMC leaves correlations with ECF close to zero (although the partial correlations do not have opposite signs). Thus, DMC is a significant, independent predictor of these important real-world behaviors.

Hypothesis 7
DMC correlates positively with measures reflecting a positive social and family environment.

Table 5's final section considers the four CEDAR social and family measures. Performance on the DMC component measures correlates negatively with father's history of drug abuse or dependence (HAR) and positively with family socioeconomic status and positive peer environment. The surprising negative correlations with social support reflect a few outliers (possibly gang members), who are low in DMC but high in self-reported social support. As with risk behaviors, these relationships are strongest for consistency in risk perception, applying decision rules, and under/overconfidence. Resistance to framing also has a statistically significant relationship here.

As would be expected from Table 5, DMC scores were strongly correlated with each of these measures (Table 6). The semipartial correlations show that DMC has independent relationships with each measure of social and family environment, after considering vocabulary and ECF, separately and jointly. The relationship with positive peer environment remains with the one-mediator null hypothesis test as well.

DMC vs. measures of cognitive ability

Table 7 reverses the order of Table 6's analyses. It looks first at simple correlations of the CEDAR measures with vocabulary and ECF, then partials out the overall DMC score. Vocabulary is little related to the CEDAR measures, other than social and family influences (and polarized thinking). ECF is related to members of each category. Generally speaking, partialling out DMC weakens these relationships, suggesting that DMC acts a mediator or simply as a more proximal variable, among correlated predictors.

CONCLUSIONS

Summary of results

Overall, these results support the external validity of conventional behavioral decision-making tasks, as represented by the seven used here. Expanding on several earlier studies, which used fewer tasks and correlates, we found that decision-making performance was predictably related to measures of (a) basic cognitive abilities, (b) cognitive styles, (c) developmental conditions, and (d) risk-taking behaviors suggesting poor

Table 7. Correlations between CEDAR variables and two measures of cognitive ability

			Semipartial correlations, controlling for DMC	
Correlated with	Vocabulary	ECF	Vocabulary	ECF
Cognitive ability				
Vocabulary	—	0.51	—	0.32
ECF	0.51	—	0.32	—
*Overall**	$p < 0.0001$		$p = 0.0001$	
Cognitive style				
Polarized thinking	−0.30	−0.27	−0.17	−0.18
Self-consciousness	0.06	0.33	−0.04	0.26
Self-monitoring	−0.03	−0.04	−0.16	−0.14
Behavioral coping	0.16	0.14	0.03	0.02
Overall	$p = 0.09$	$p = 0.005$	$p = 0.25$	$p = 0.03$
Risk behavior				
Antisocial disorders	−0.08	−0.33	0.01	−0.24
Externalizing behavior	−0.15	−0.41	0.01	−0.27
Delinquency	−0.10	−0.31	0.05	−0.19
l*n*(lifetime # of drinks)	0.02	−0.13	0.10	−0.04
l*n*(lifetime marijuana use)	0.03	−0.17	0.15	−0.05
l*n*(# times had sex)	0.06	−0.14	0.18	−0.04
l*n*(# sexual partners)	−0.01	−0.13	0.13	−0.01
Overall	$p = 0.64$	$p = 0.004$	$p = 0.38$	$p = 0.19$
Social and family influences				
Risk status (HAR = 1; LAR = 0)	−0.23	−0.31	−0.06	−0.14
SES	0.36	0.31	0.20	0.16
Social support	−0.28	−0.21	−0.15	−0.07
Positive peer environment	0.02	0.16	−0.13	0.03
Overall	$p = 0.0007$	$p < 0.0001$	$p = 0.03$	$p = 0.16$

*Overall p-values were computed using Strube's (1985) method for combining significance levels from nonindependent hypothesis tests. All reported ps are two-sided. The correlation between CEDAR measures was used as an estimate of the correlation between test statistics. A conservative Bonferroni correction on the 57 tests presented here and in Tables 5 and 6 would convert an individual $\alpha = 0.05$ into an $\alpha = 0.0009$. If one were interested in an individual zero-order correlation, approximate cutoffs would be $r = 0.16$, $p < 0.05$; $r = 0.22$, $p < 0.01$; $r = 0.29$, $p < 0.001$. For a semipartial correlation, approximate cutoffs would be $r = 0.18$, $p < 0.05$; $r = 0.25$, $p < 0.01$; $r = 0.32$, $p < 0.001$.

real-world decision making. These relationships were found for performance on individual decision-making tasks and for DMC factor scores derived from a single-factor model. The three measures loading highly on that factor cut across the three distinctions underlying our tasks (Table 1), suggesting that it captures some global competence.

The strong correlations with vocabulary and ECF support DMC being a form of cognitive competence. The strong correlations with the cognitive-style measures support DMC being associated with constructive and introspective thinking. The strong correlations with developmental environment (e.g., risk status, positive peer environment, SES) are consistent with DMC arising in what should be conducive conditions. The strong correlations with risk behaviors (delinquency, marijuana use, and early sexual behavior) suggest that DMC captures abilities that matter in everyday life, for individuals like those in the CEDAR sample. Such external validation emerges even though the prototypes for our DMC component tasks were developed and refined primarily by investigators studying general cognitive processes, typically in experiments with college students.

The analyses in Tables 6 and 7 indicate that the DMC tasks are more than just a surrogate for cognitive ability. Controlling for vocabulary and ECF left the relationships largely unchanged, with both semipartial correlation analyses and tests of a one-mediator null hypothesis—although the latter showed some role for ECF.

Of course, correlation need not imply causality. Risk behaviors could be both cause and effect of poor decision making. People who have trouble controlling themselves may not stop to think; people who experience difficulty making decisions may get frustrated, then act out. A negative environment could similarly work both ways. Delinquent friends may degrade decision making, while poor decision making could increase the chance of having deviant peers. The two measures of family circumstances should represent just causes of poor DMC. Growing up in a troubled household could reduce DMC, by depriving young people of role models and disturbing their learning (not to mention any genetic link). For other relationships, a more elaborate research design is needed to discern cause and effect.

The positive performance manifold in the decision-making tasks supports Stanovich and West's (1998, 2000) prediction of its existence, and their associated claim that judgmental biases are more than just random performance errors. DMC's correlation with other cognitive abilities suggests that cognitive limits might restrict decision-making performance. Stanovich and West proposed that people sometimes perform poorly because they deliberately use a non-normative strategy, not recognizing its inappropriateness. That tendency might be reflected in the strong positive correlation between DMC and behavioral coping (measuring endorsement of normative decision-making strategies).

Our study extends those of Stanovich, West, Levin and others by (a) offering a more systematic approach to selecting tasks; (b) using a more comprehensive set of tasks, some not previously addressed in such studies; (c) expanding the set of covariates, beyond cognitive ability and cognitive styles, to include demographic antecedents and real-world behavior; and (d) considering the possibility of an overall DMC measure. This last issue builds on Stanovich and West's proposal of a "positive manifold," providing broader evidence regarding the potential usefulness and external validity of such a measure. Our results also suggest the need for better understanding of component DMC skills, beyond that central factor.

Potential limitations

Sample

One potentially relevant feature of the CEDAR sample is that all participants were male. Tests of mathematical and verbal cognitive abilities have not shown consistent gender differences of any noteworthy magnitude (Cole, 1997; Hedges & Nowell, 1995; Jensen, 1998). The few gender differences found in decision-making tasks have typically involved values, rather than performance. For example, Eckel and Grossman found that women cooperate more in some ultimatum games (1998) and care more about fairness (1996). Larrick et al. (1993) found that women are less willing than men to use cost-benefit reasoning when such reasoning conflicts with social values, but are equally likely to use such reasoning when no such conflict exists. Gilligan (1982) argued that males are more logic driven and females more relationship focused on moral reasoning tasks. Such differences might, conceivably, lead females to have lower scores on our recognizing social norms, applying decision rules, and value-assessment tasks. Even if that were the case, it need not affect relationships between DMC scores and the CEDAR variables—although correlations might be reduced by women's restricted range of variation on many risk behaviors (e.g., antisocial disorders, externalizing, delinquency). Donovan and Jessor (1985) found that adolescent problem behaviors form a single factor, but one that it is more internally consistent for males than for females. Replication with female subjects is clearly an important topic for future research. Considering these related results, we would predict similar, but perhaps weaker patterns.

A second, potentially relevant, feature of the CEDAR sample is its having deliberately oversampled high-risk youth and, hence, that tail of the distributions for risk-related measures. If HAR and LAR respondents had different patterns with respect to DMC, the correlations in any overall sample would reflect its mix of HAR and LAR respondents. At the extreme, the correlations with the continuous variables could reflect just the between-group differences. However, separate analyses on the LAR ($n = 69$) and HAR ($n = 41$) subsamples revealed similar first-order correlations between DMC and other variables, although typically stronger with LAR teens. There was only one statistically significant difference ($\alpha = 0.05$; two-tailed test) between correlations in the two subsamples: polarized thinking was more strongly correlated to DMC for LAR respondents (-0.49 vs. -0.03; $z = 2.25$, $p < 0.05$).[10]

Finally, the limited size of the available CEDAR sample ($N = 110$) restricted the opportunities for more detailed analyses. A larger sample would allow more precise estimates and more complex analyses, such as concurrent structural equation models of the DMC factor structure and the two measures of cognitive ability.

Tasks

Our tasks were based on ones commonly used in the research literature, seeking to represent each of the four core skill areas, including both judgment and choice tests, with both accuracy and consistency criteria. As mentioned, other tasks could be chosen for each skill. For example, belief assessment could also be measured by tasks requiring respondents to judge randomness or to combine individuating and base-rate information. More fully assessing the role of subdomains will require multiple tasks representing each. However, as seen here, it is often difficult to assign tasks to a single domain (possibly limiting the opportunities to test Hypothesis 3). Another theoretical consideration in choosing tasks is whether to represent the skill in general terms or a specific domain. We pursued the former strategy. Grisso et al. (2003), also Scott et al. (1995) sought a specific measure of adolescents' competence for adjudication. Future work might also consider the extent to which tasks tap into different processing systems, as captured in dual-process theories (e.g., Kirkpatrick & Epstein, 1992).

Within these theoretical constraints, a measure should have attractive psychometric properties. We, in effect, gambled that our seven tasks, chosen for their face validity in representing the domains, would show enough internal and external validity to merit further research. We believe that this study, the first of its kind, showed sufficiently promising results to warrant systematic development of the constituent tasks, leading to a standardized measure of DMC. Three of our tasks showed limited score variance and internal validity (as represented by low Cronbach's alpha): resistance to sunk cost, resistance to framing, and consistency in risk perception. Some of this poor psychometric performance may reflect the small numbers of items and limited response scales for these tasks, limiting their ability to detect latent covariation. Some, though, may reflect the tasks capturing more than one underlying phenomenon—which might be individually important. For example, resistance to framing showed little internal consistency, yet still correlated as expected with many of the CEDAR covariates. Further research, with more systematic task development, is needed to clarify these possibilities. The logical next step in that research is increasing score variance. If that does not improve internal validity, then alternative representatives of those skills are needed, before reaching any conclusions about their external validity. Additional task development should also improve the modest test–retest reliability ($r = 0.54$). Once completed, it would allow sounder answers regarding the factor structure of DMC and the real-world correlates of particular skills.

One possible extension of the present task set is to add tasks that assess skill in decision structuring. These could examine individuals' ability to identify relevant (a) courses of action, (b) potential consequences of those actions, (c) sources of uncertainty (regarding which consequences will follow each action), and (d) rules for integrating decision-relevant information. The present tasks all require exercising skills within already created structures. Decision structuring requires open-ended tasks, which were beyond the technical constraints of administration of the CEDAR sample. Some candidate tasks appear in Beyth-Marom et al.

[10]Another possible difference is in recruitment, whether at CEDAR or by mail. HAR were similarly represented in the two groups (39.3% and 34.7%, respectively; $\chi^2(1) = 0.25$, ns.). The two subsamples were similar in SES ($t(108) = -1.57$, ns.). However, mail-in respondents had higher vocabulary ($t(105) = -2.88$, $p = 0.005$), ECF ($t(108) = -4.65$, $p < 0.001$), and DMC scores ($t(107) = -2.27$, $p = 0.025$). Repeating the analyses for respondents recruited each way yielded similar results in the each subsample (recognizing that the reduced sample size provided statistical power for observing only relatively large differences).

(1993), Fischhoff (1992, 1996), Fischhoff et al. (1998), Hogarth and Kunreuther (1995), and Jones et al. (1997).

Stronger component measures will also clarify the meaning of the weak inter-correlations among the DMC component tasks. They could mean that any omnibus DMC measure can be usefully supplemented by assessment of individual skills. However, the relatively strong nomological validity does support some significant role for a central skill within DMC. As mentioned, the overall DMC measure is related to an impressive number of covariates, with many of these relationships being independent of two well-regarded general measures of cognitive ability, vocabulary and ECF. This makes sense, as overall DMC, a weighted average, will naturally be less susceptible to the random noise component of unreliability. The test–retest reliability of the DMC scores was 0.54, large enough to indicate a stable trait, but small enough to set an upper bond on validity. Thus, even with this imperfect assessment of DMC, the present results support the external validity of the sort of tasks commonly used in behavioral decision-making research, as relevant to real-world phenomena.

The content and temporal ordering of the CEDAR covariates, designed well before the present study, provide a final limitation to the present results. WISC-R was only run at the baseline (age 10–12) assessment; other measures were gathered at age 16—before the present assessment, at ages 18–19. In these cases DMC's predictive ability, relative to the cognitive ability variables, may benefit from a temporal-proximity advantage, to the extent that cognitive ability changes over adolescence. Cognitive style measures that would be interesting additions in future research include need for cognition (Cacioppo & Petty, 1982), need for closure (Webster & Kruglanski, 1994), and Schwartz et al.'s (2002) measure of maximizing vs. satisficing.

Our seven-task DMC battery took 30–45 minutes to administer. Researchers interested in examining its covariates, but with less time, might focus on the three tasks that showed the greatest explanatory power: under/overconfidence, applying decision rules, and resistance to framing. Of these, the confidence task has the best psychometric properties, reflecting some mixture of the task's structure (having the most items and, perhaps, the longest experimental pedigree) and theoretical importance. Better component tasks and understanding of their covariates should contribute to better understanding of overall DMC, as has happened with scientific understanding of intelligence.

Relationship to intelligence and real-life consequents

A natural question is the relationship between DMC and the more general cognitive ability commonly called "intelligence." Jensen (1998) seconds Spearman in questioning the scientific usefulness of the term, arguing that it incorporates such diverse concepts and connotations as to preclude agreement on a precise scientific definition. Thus, although vocabulary and ECF load heavily on the two key concepts of crystallized (*Gc*) and fluid (*Gf*) intelligence, one might still argue that our controlling for them did not sufficiently account for the influence of "intelligence." Jensen proposes, instead, focusing on specific mental abilities, which can be defined more closely and might reflect a shared underlying factor (g). DMC might be one such ability.

Given its strong correlations with vocabulary and ECF, DMC should be *g*-loaded as well. Its general value would come from having independent or superior predictive value, relative to other indicators of *g*. In terms of zero-order correlations, DMC predicted the CEDAR covariates better than either vocabulary or ECF. It also had incremental predictive power, as shown in the semipartial and one-mediator null hypothesis analyses. Despite these suggestive results, however, the present data permit only limited tests of discriminant validity. A full multitrait-multimethod study (Campbell & Fiske, 1959) is needed to distinguish properly between DMC and other cognitive abilities. However, even if DMC represents "just" a combination of general cognitive abilities, it is a combination that seems to provide a potentially useful measure of competence—one that might serve as a predictor, consequence, or correlate of important real-world behaviors.

A related question is whether better decision making in the laboratory indeed has an influence on important real-world outcomes. Larrick et al. (1993) found evidence that greater consistency with cost-benefit rules corresponds to both higher cognitive ability and greater eventual salary. Although the data are associative in nature, the authors argue that both relationships are necessary for the external validity of cost-benefit reasoning tasks. A similar argument can be made more generally with many behavioral decision-research tasks, such as those considered here. We provide preliminary (correlative) evidence that greater general cognitive ability is related to greater DMC, as well as many real-world outcomes relevant to the risks faced by teens. In principle, however, if such phenomena as framing effects and overconfidence (as observed in the laboratory) are ecologically valid, we should be able to measure their impact on achievement of goals in life, relative to the opportunities that individuals have. An additional issue not addressed by Larrick et al. (1993) is whether better decision making is actually a stronger predictor of such real-world success than is general cognitive ability (e.g., through mediation). The semipartial analyses presented here suggest that this may be the case.

Final remarks

Within these constraints, our results provide preliminary evidence for DMC's validity as a psychological construct (Cronbach & Meehl, 1955; Nunnally & Bernstein, 1994). The individual DMC measures formed a moderate positive manifold, demonstrating some convergent validity. They also showed some nomological validity, in the correlations between performance on DMC tasks and diverse measures of cognitive ability, cognitive style, risk behavior, and social environment. These correlations were even more pronounced when using a weighted average of the seven individual measures (DMC scores). They remained after controlling for well-regarded measures of intelligence. They seem to present strongly suggestive evidence for the external validity of conventional decision-making tasks and reason for additional research, more fully examining the strength, direction, and practical importance of these relationships.

APPENDIX

The *one-mediator null hypothesis* (Birnbaum, 1985) provides an alternative to partial (or semipartial) correlations for testing the link between DMC and another measure, while accounting for cognitive ability. Figure A1 depicts this hypothesis, using SES as an example. DMC, ECF, and SES are all observed variables, while cognitive ability is an unobserved latent trait. The three *e*'s are mutually independent error terms, and

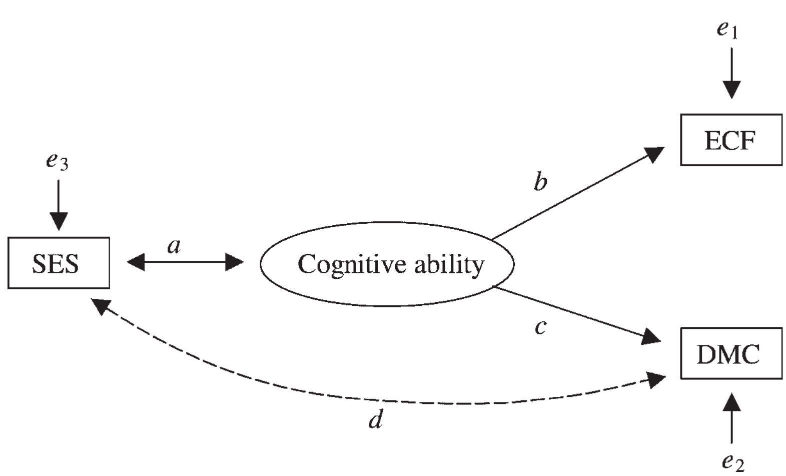


Figure A1. Example of the one-mediator null hypothesis: d = 0?

a, *b*, *c*, and *d* are correlations. The one-mediator null hypothesis states that DMC and ECF (or, alternatively, vocabulary) are both imperfect measures of the underlying construct of cognitive ability, which is, in fact, the variable related to SES. If cognitive ability completely mediates the relationship between DMC and SES, then d would equal 0. The alternative hypothesis posits a direct link between DMC and SES (i.e., $d \neq 0$).

Two tests of whether $d = 0$ have been proposed. Test A is based on the observation that, if the null hypothesis holds, then:

$$\rho_1 \leq \frac{\rho_3}{\rho_2} \leq \frac{1}{\rho_1},$$

where

$$\begin{aligned}
\rho_1 &= \rho(\mathrm{DMC},\ \mathrm{ECF}) = \mathrm{bc} + \mathrm{abd}, \\
\rho_2 &= \rho(\mathrm{ECF},\ \mathrm{SES}) = \mathrm{ab},\ \text{and} \\
\rho_3 &= \rho(\mathrm{DMC},\ \mathrm{SES}) = \mathrm{ac} + \mathrm{d}.
\end{aligned}$$

That is, the ratio of the correlation between DMC and SES to the correlation between ECF and SES must be larger than the correlation between DMC and ECF *and* smaller than its reciprocal. Unless this condition holds, we have evidence against the one-mediator null hypothesis (and, consequently, for a direct link between DMC and SES). This test is generally considered to be quite conservative (Birnbaum, 1981; McLaughlin, 1982).

Test B is even more conservative (although not strictly nested in Test A). It requires rejecting the null hypothesis, if the sign of the partial correlation between DMC and SES, controlling for ECF, is the opposite of the sign of the partial correlation between ECF and SES, controlling for DMC. If so, then there clearly cannot be (complete) mediation. Therefore, Tests A and B provide progressively stronger evidence against the one-mediator null hypothesis.

In the second and third columns of Table 5, superscripts *a* and *b* represent cases where Tests A and B, respectively, suggest rejecting the one-mediator null hypothesis. For brevity, only the more stringent test that rejects the null hypothesis is listed (i.e., *b* usually implies *a*, as well).

ACKNOWLEDGEMENTS

The Center for Education and Drug Abuse Research is an ongoing study conducted at the School of Pharmacy, University of Pittsburgh, supported by the National Institute on Drug Abuse. This study is also supported by grants from the National Institute of Allergies and Infectious Diseases and the National Science Foundation (SBR-9521914, SES-0213782).

The authors thank Peter Giancola, Irwin Levin, and for Ralph Tarter their contributions and support. We also thank Robyn Dawes, Paul Fischbeck, Wändi Bruine de Bruin, Julie Downs, Claire Palmgren, and the rest of the adolescent-health research group for their invaluable criticisms, suggestions, and insights—as well as several anonymous reviewers.

REFERENCES

Achenback, T. M., & Edelbrock, C. (1983). *Manual for the child behavior checklist and revised child behavior profile*. Burlington, VT: T. M. Achenback.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*, 124–140.
Baron, J. (1985). *Rationality and intelligence*. New York: Cambridge University Press.
Baron, J. (1988). *Thinking and deciding*. New York: Cambridge University Press.
Baron, J., Granato, L., Spranca, M., & Teubal, E. (1993). Decision-making biases in children and early adolescents: Exploratory studies. *Merrill-Palmer Quarterly, 39*, 22–46.
Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*, 569–579.
Beyth-Marom, R., Austin, L., Fischhoff, B., Palmgren, C., & Quadrel, M. J. (1993). Perceived consequences of risky behaviors. *Developmental Psychology, 29*, 549–563.
Birnbaum, M. H. (1981). Reply to McLaughlin: Proper path models for theoretical partialing. *American Psychologist, 36*, 1193–1195.
Birnbaum, M. H. (1985). Relationships among models of salary bias. *American Psychologist, 40*, 862–866.
Bromiley, P., & Curley, S. P. (1992). Individual differences in risk taking. In J. F. Yates (Ed.), *Risk-taking behavior* (pp. 87–132). New York, NY: John Wiley & Sons.
Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.
Camerer, C. F. (1992). Recent tests of generalizations of expected utility theory. In W. Edwards (Ed.), *Utility theories: Measurements and applications* (pp. 207–251). Boston: Kluwer Academic Publishers.
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.
Cohen, S., & Hoberman, H. M. (1983). Positive events and social supports as buffers of life change stress. *Journal of Applied Social Psychology, 13*, 99–125.
Cole, N. S. (1997). *The ETS gender study: How females and males perform in educational settings*. Princeton, NJ: Educational Testing Service.
Cole, P., Usher, B., & Cargo, A. (1993). Cognitive risk and its association with risk for disruptive behavior disorder in preschoolers. *Journal of Clinical Child Psychology, 22*, 154–164.
Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104.
Crawford, J. D., & Stankov, L. (1996). Age differences in the realism of confidence judgments: a calibration study using tests of fluid and crystallized intelligence. *Learning and Individual Differences, 8*, 83–103.
Cronbach, L. J. (1951). Coefficient alpha and the internal structure of the tests. *Psychometrika, 16*, 297–334.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
Dawes, R. M. (1990). The potential nonfalsity of the false consensus effect. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn*. Chicago, IL: University of Chicago Press.
Dawes, R. M., & Hastie, R. (2001). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Thousand Oaks, CA: Sage Publications.
Donovan, J. E., & Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. *Journal of Consulting and Clinical Psychology, 53*, 890–904.
Dryfoos, J. G. (1990). *Adolescents at risk*. New York, NY: Oxford University Press.
Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183–201.
Eckel, C. C., & Grossman, P. J. (1996). The relative price of fairness: gender differences in a punishment game. *Journal of Economic Behavior and Organization, 30*, 143–158.
Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men? Evidence from dictator experiments. *The Economic Journal, 108*, 726–735.
Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380–417.
Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*, 193–242.
Endicott, J., & Spitzer, R.L. (1978). A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of General Psychiatry, 35*, 837–844.
Epstein, S., & Meier, P. (1989). Constructive thinking: a broad coping variable with specific components. *Journal of Personality and Social Psychology, 57*, 332–350.
Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522–527.

Finucane, M. L., Slovic, P., Hibbard, J. H., Peters, E., Mertz, C. K., & Macgregor, D. G. (2002). Aging and decision-making competence: an analysis of comprehension and consistency skills in older versus younger adults considering health-plan options. *Journal of Behavioral Decision Making, 15*, 141–164.
Fischhoff, B. (1983). Predicting frames. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 113–116.
Fischhoff, B. (1991). Value elicitation: is there anything in there? *American Psychologist, 46*, 835–847.
Fischhoff, B. (1992). Giving advice: decision theory perspectives on sexual assault. *American Psychologist, 47*, 577–588.
Fischhoff, B. (1993). Transaction analysis: a framework and an application to insurance decisions. *Journal of Risk and Uncertainty, 7*, 53–69.
Fischhoff, B. (1996). The real world: what good is it? *Organizational Behavior and Human Decision Processes, 65*, 232–248.
Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90*, 239–260.
Fischhoff, B., Bostrom, A., & Quadrel, M. J. (2002). Risk perception and communication. In R. Detels, J. McEwen, R. Beaglehole, & H. Tanaka (Eds.), *Oxford textbook of public health.* London: Oxford University Press.
Fischhoff, B., Downs, J. S., & Bruine de Bruin, W. (1998). Adolescent vulnerability: a framework for behavioral interventions. *Applied and Preventive Psychology, 7*, 77–94.
Fischhoff, B., & MacGregor, D. (1983). Judged lethality: how much people seem to know depends upon how they are asked. *Risk Analysis, 3*, 229–236.
Fischhoff, B., Parker, A. M., Bruine de Bruin, W., Downs, J. S., Palmgren, C., Dawes, R. M., & Manski, C. (2000). Teen expectations for significant life events. *Public Opinion Quarterly, 64*, 189–205.
Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
Freund, L. S. (1990). Maternal regulation of children's problem-solving behavior and its impact on children's performance. *Child Development, 61*, 113–126.
Frisch, D. (1993). Reasons for framing effects. *Organizational Behavior and Human Decision Processes, 54*, 399–429.
Giancola, P. R., Martin, C. S., Tarter, R. E., Pelham, W. E., & Moss, H. B. (1996). Executive cognitive functioning and aggressive behavior in preadolescent boys at high risk for substance abuse/dependence. *Journal of Studies on Alcohol, 57*, 352–359.
Giancola, P. R., Mezzich, A. C., & Tarter, R. E. (1998). Executive cognitive functioning, temperament, and antisocial behavior in conduct-disordered adolescent females. *Journal of Abnormal Psychology, 107*, 629–641.
Gigerenzer, G., Todd, P. M., & the ABC Group (2000). *Simple heuristics that make us smart*. Cary, NC: Oxford University Press.
Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
Graziano, W. G., Leone, C., Musser, L. M., & Lautenschlager, G. J. (1987). Self-monitoring in children: a differential approach to social development. *Developmental Psychology, 23*, 571–576.
Grisso, T., Steinberg, L., Woolard, J., Cauffman, E., Scott, E., Graham, S., Lexcen, F., Reppucci, N. D., & Schwartz, R. (2003). Juveniles' competence to stand trial: A comparison of adolescents' and adults' capacities as trial defendants. *Law and Human Behavior, 27*, 333–363.
Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41–45.
Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology, 53*, 221–234.
Hogarth, R. M., & Kunreuther, H. (1995). Decision making under ignorance: Arguing with yourself. *Journal of Risk and Uncertainty, 10*, 15–36.
Hollingshead, A. (1975). *Four factor index of social status*. University of Connecticut: Unpublished manuscript.
Institute of Medicine. (1999). *Adolescent decision making: Implications for prevention programs*. Washington, DC: National Academy Press.
Jacobs, J. E., Greenwald, J. P., & Osgood, D. W. (1995). Developmental differences in baserate estimates of social behaviors and attitudes. *Social Development, 4*, 165–181.
Janis, I. L., & Mann, L. (1976). Coping with decisional conflict: An analysis of how stress affects decision-making suggests interventions to improve the process. *American Scientist, 64*, 657–666.
Jensen, A. R. (1998). *The* g *factor.* Westport, CT: Praeger.
Jessor, R., & Jessor, S. (1977). *Problem behavior and psychopathology.* New York, NY: Academic Press.

Jones, S. K., Yurak, T. J., & Frisch, D. (1997). The effect of outcome information on the evaluation and recall of individuals' own decisions. *Organizational Behavior and Human Decision Processes, 71*, 95–120.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.
Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*, 263–291.
Katz, L., & Epstein, S. (1991). Constructive thinking and coping with laboratory-induced stress. *Journal of Personality and Social Psychology, 61*, 789–800.
Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 24*, 411–435.
Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social Psychology, 63*, 534–544.
Klein, G. (1999). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.
Larrick, R. P., Nisbett, R. E., & Morgan, J. N. (1993). Who uses the cost-benefit rules of choice? Implications for the normative status of microeconomic theory. *Organization Behavior and Human Decision Processes, 56*, 331–347.
Levin, I. P. (1999, November). *Why do you and I make different decisions? Tracking individual differences in decision making*. Presidential address to the Society for Judgment and Decision Making, Los Angeles, CA.
Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: need for cognition and inclusion–exclusion effects. *Organizational Behavior and Human Decision Processes, 82*, 171–193.
Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*, 149–188.
Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance, 20*, 159–183.
Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 551–578.
Linville, P. W., Fischer, G. W., & Fischhoff, B. (1993). AIDS risk perceptions and decision biases. In J. B. Pryor & G. D. Reeder (Eds.), *The social psychology of HIV infection*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Loeber, R. (1989). *Values and attitudes towards delinquent behavior*. Pittsburgh Youth Study. Department of Psychiatry, University of Pittsburgh, PA.
Lopes, L. L. (1987). Between hope and fear: the psychology of risk. *Advances in Experimental Social Psychology, 20*, 255–295.
Luce, M. F. (1998). Choosing to avoid: coping with negatively emotion-laden consumer decisions. *Journal of Consumer Research, 24*, 409–433.
Mano, H. (1992). Judgments under distress: assessing the role of unpleasantness and arousal in judgment formation. *Organizational Behavior and Human Decision Processes, 52*, 216–245.
Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.
McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive *Gf–Gc* framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 151–179). New York: Guilford Press.
McLaughlin, S. D. (1982). Reply to Birnbaum. *American Psychologist, 37*, 1142–1144.
Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology, 53*, 298–305.
Murphy, R. O. (1997, November). *Stability of calibration biases over time*. Poster presented at the Annual Meeting of the Society for Judgment and Decision Making, Philadelphia, PA.
National Longitudinal Study of Youth. (1997). Bureau of Labor Statistics, US Department of Labor.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
Overton, W. F. (1990). Competence and procedures: Constraints on the development of logical reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic* (pp. 1–32). Hillsdale, NJ: Lawrence Erlbaum, Associates.
Pallier, G., Wilkinson, R., Danthir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology, 129*, 257–299.
Parker, A. M., Downs, J. S., Fischhoff, B., Bruine de Bruin, W., & Dawes, R. M. (2000). *The appropriateness of adolescents' confidence in their knowledge: AIDS-related and general*. Carnegie Mellon University: Unpublished Manuscript.
Parrill-Burnstein, M. (1978). Teaching kindergarten children to solve problems: an information-processing approach. *Child Development, 49*, 700–706.

Parsons, O., Tarter, R., & Edelberg, R. (1972). Altered motor control in chronic alcoholics. *Journal of Abnormal Psychology, 80*, 308–314.
Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.
Pelham, W., Schneider, W., Evans, S., & Gnagy, E. (1987). *Sustained attention and ADHD: Vigilance performance and methylphenidate effects in the laboratory and the natural environment*. Paper presented at the Society for Research on Child and Adolescent Psychopathology, Sarasota, FL.
Peterson, C. R., Ulehla, Z. J., Miller, A. J., & Bourne, Jr., L. E. (1965). Internal consistency of subjective probabilities. *Journal of Experimental Psychology, 70*, 526–533.
Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.
Porteus, S. (1965). *Porteus maze test: Fifty year's application*. Palo Alto, CA: Pacific Books.
Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology, 64*, 243–256.
Raiffa, H. (1968). *Decision analysis*. Reading, MA: Addison-Wesley.
Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shaw, M., Ireland, M., Bearinger, L. H., & Udry, J. R. (1997). Protecting adolescents from harm: Findings from the National Longitudinal Study on Adolescent Health. *Journal of the American Medical Association, 278*, 823–832.
Roelofsma, P. H. M. P., & Keren, G. (1995). Framing and time-inconsistent preferences. In J.-P. Caverni, M. Bar-Hillel, F. H. Barron, & H. Jungermann (Eds.), *Contributions to decision making–I*. New York, NY: Elsevier Science BV.
Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence, and limitations. *Journal of Economic Literature, 20*, 529–563.
Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology, 83*, 1178–1197.
Scott, E. S., Reppucci, N. D., & Woolard, J. L. (1995). Evaluating adolescent decision making in legal contexts. *Law and Human Behavior, 19*, 221–344.
Shafir, E. (1993). Choosing versus rejecting: why some options are both better and worse than others. *Memory and Cognition, 21*, 546–556.
Silverman, I., & Ragusa, D. (1992). A short-term longitudinal study of the early development of self-regulation. *Journal of Abnormal Child Psychology, 20*, 415–435.
Skinner, H. A. (1982). *Development and validation of lifetime alcohol consumption assessment procedure*. Substudy No. 1284. Toronto: Addiction Research Foundation.
Slovic, P. (1962). Convergent validation of risk-taking measures. *Journal of Abnormal and Social Psychology, 65*, 68–71.
Slovic, P (Ed.). (2001). *Perception of risk*. London: Earthscan.
Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making, 9*, 283–290.
Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*, 526–537.
Snyder, M., & Cantor, N. (1980). Thinking about ourselves and others: self-monitoring and social knowledge. *Journal of Personality and Social Psychology, 39*, 222–234.
Spitzer, R. L., Williams, B. W., & Gibbon, M. (1987). *Instruction manual for the structured clinical interview for DSM-III-R*. Biometrics Research Department. New York, NY: New York State Psychiatric Institute.
Stankov, L. (1998). Calibration curves, scatterplots, and the distinction between general knowledge and perceptual tests. *Learning and Individual Differences, 10*, 29–50.
Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences, 21*, 971–986.
Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence, 25*, 93–109.
Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161–188.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645–726.
Strube, M. J. (1985). Combining and comparing significance levels from nonindependent hypothesis tests. *Psychological Bulletin, 97*, 334–341.
Tarter, R. E. (1991). *Child report on peer environment*. Unpublished instrument, Center for Education and Drug Abuse Research, Pittsburgh, PA.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.

Tversky, A., & Kahneman, D. (1988). Rational choice and the framing of decisions. In D. E. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions*. New York: Cambridge University Press.
Vlek, C., & Stallen, P.-J. (1981). Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance, 28*, 235–271.
Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: a psychological and psychometric review. *Management Science, 29*, 151–173.
Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*, 263–290.
Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Personality Processes and Individual Differences, 67*, 1049–1062.
Wechsler, D. (1972). *Wechsler Intelligence Scale for Children–Revised*. New York: The Psychological Corporation.
Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised*. New York: The Psychological Corporation.
West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: individual difference in performance estimation bias. *Psychonomic Bulletin and Review, 4*, 387–392.
Whitcomb, K. M., Önkal, D., Benson, P. G., & Curley, S. P. (1993). An evaluation of the reliability of probability judgments across response modes and over time. *Journal of Behavioral Decision Making, 6*, 283–296.
Wolfe, R. N., & Grosch, J. W. (1990). Personality correlates of confidence in one's decisions. *Journal of Personality, 58*, 515–534.
Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice Hall.
Yates, J. F. (Ed.). (1992). *Risk-taking behavior*. Chichester: John Wiley & Sons.

Authors' biographies:

Andrew M. Parker is an assistant professor in the Department of Marketing at Virginia Polytechnic Institute and State University. His research includes individual differences in decision making, risk perception and communication, and confidence in judgment.

Baruch Fischhoff is University Professor, in the Department of Social and Decision Sciences and Department of Engineering and Public Policy at Carnegie Mellon University, and a member of the Institute of Medicine. His research includes risk perception and communication, risk management, adolescent decision making, medical informed consent, and environmental protection.

Authors' addresses:

Andrew M. Parker, Department of Marketing (0236), Virginia Tech, Blacksburg, VA 24061, USA.

Baruch Fischhoff, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

