

Reasoning Independently of Prior Belief and Individual Differences in Actively Open-Minded Thinking

Keith E. Stanovich
University of Toronto

Richard F. West
James Madison University

A sample of 349 college students completed an argument evaluation test (AET) in which they evaluated arguments concerning real-life situations. A separate regression analysis was conducted for each student predicting his or her evaluations of argument quality from an objective indicator of argument quality and the strength of his or her prior beliefs about the target propositions. The beta weight for objective argument quality was interpreted in this analysis as an indicator of the ability to evaluate objective argument quality independent of prior belief. Individual differences in this index were reliably linked to individual differences in cognitive ability and actively open-minded thinking dispositions. Further, actively open-minded thinking predicted variance in AET performance even after individual differences in cognitive ability had been partialled out.

Discussions of critical thinking in the educational and psychological literature consistently point to the importance of decontextualized reasoning styles that foster the tendency to evaluate arguments and evidence in a way that is not contaminated by one's prior beliefs. The disposition toward such unbiased reasoning is almost universally viewed as a characteristic of good critical thought. For example, Norris and Ennis (1989) listed as one characteristic of critical thinkers the disposition to "reason from starting points with which they disagree without letting the disagreement interfere with their reasoning" (p. 12). Zechmeister and Johnson (1992) listed as one characteristic of the critical thinker the ability to "accept statements as true even when they don't agree with one's own position" (p. 6). Similarly, Nickerson (1987) and many other theorists (e.g., Brookfield, 1987; Lipman, 1991; Perkins, 1995; Perkins, Jay, & Tishman, 1993; Wade & Tavris, 1993) stressed that critical thinking entails the ability to recognize "the fallibility of one's own opinions, the probability of bias in those opinions, and the danger of differentially weighting evidence according to personal preferences" (Nickerson, 1987, p. 30). The grow-

ing literature on informal or practical reasoning likewise emphasizes the importance of detaching one's own beliefs from the process of argument evaluation (Baron, 1991, 1995; Brenner, Koehler, & Tversky, 1996; Galotti, 1989; Kardash & Scholes, 1996; Klaczynski, 1997; Klaczynski & Gordon, 1996; Kuhn, 1991, 1993; Voss, Perkins, & Segal, 1991).

The emphasis on avoiding the effects of belief bias in the critical thinking literature is well motivated by the theoretical prominence accorded the concept of cognitive decontextualization within psychology (e.g., Donaldson, 1978, 1993; Epstein, 1994; Evans, 1984, 1989; Luria, 1976; Neimark, 1987; Olson, 1977, 1986, 1994; Piaget, 1926, 1972). An important research tradition within cognitive science has examined the influence of prior beliefs on argument evaluation and has demonstrated how prior belief does bias human reasoning (Baron, 1995; Broniarczyk & Alba, 1994; Evans, Over, & Manktelow, 1993; Markovits & Nantel, 1989; Oakhill, Johnson-Laird, & Garnham, 1989). Although investigators have observed group trends indicating belief-bias effects in the literatures of both social psychology (e.g., Lord, Lepper, & Preston, 1984) and cognitive psychology (Evans, Barston, & Pollard, 1983; Oakhill et al., 1989), only recently have they turned their attention to individual differences in these effects (see Kardash & Scholes, 1996; Klaczynski, 1997; Kuhn, 1991, 1993; Slusher & Anderson, 1996).

The relative lack of work on individual differences in reasoning in the presence of prior belief has had negative practical and theoretical consequences. On the practical side, little progress has been made in assessing this foundational aspect of critical thinking. As mentioned previously, reasoning objectively about data and arguments that contradict prior beliefs is often seen as the quintessence of critical thought. It is thus essential that one has experimental methods that can identify individual differences in this crucial component of rational thought and that can relate these differences to other cognitive and personality vari-

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

This research was supported by Grant 410-95-0315 from the Social Sciences and Humanities Research Council of Canada and a James Madison University Program Faculty Assistance grant. We thank Penny Chiappe, Alexandra Gottardo, Walter Sa, and Ron Stringer for their assistance in data coding. Anne E. Cunningham is thanked for her help in obtaining the experts' evaluations. We also thank Carole Beal for her extensive comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Keith E. Stanovich, Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education, University of Toronto, 252 Bloor Street West, Toronto, Ontario, Canada M5S 1V6.

ables. Only then will researchers know if the idea of assessing such a component of thought is even feasible. For example, many standardized critical thinking tests finesse the issue with neutral content (e.g., Level X of the *Cornell Critical Thinking Test*, Ennis & Millman, 1985). Other tests have controversial content but do not actually measure the prior belief about the content; instead they try to counterbalance prior belief so that it balances out across individuals (e.g., *Watson-Glaser Critical Thinking Appraisal*, Watson & Glaser, 1980).

The relative lack of data on individual differences in reasoning in the presence of prior belief also may have hindered efforts to evaluate the normative status (see Baron, 1985; Bell, Raiffa, & Tversky, 1988) of this thinking skill. Although, as previously mentioned, the critical thinking literature emphasizes the ability to separate prior belief from argument evaluation, there are alternative traditions in developmental and cognitive psychology from which maximizing the decontextualization of the evaluation process might not be viewed as optimal. For example, the contextualist tradition within developmental psychology emphasizes that the optimum exercise of cognitive skills often occurs in highly contextualized situations (Bronfenbrenner & Ceci, 1994; Brown, Collins, & Duguid, 1989; Ceci, 1990, 1993; Lave & Wenger, 1991; Rogoff & Lave, 1984). This framework might lead one to question whether detaching prior belief from evidence evaluation really would be the strategy of the most cognitively able participants.

Several philosophical analyses have also questioned the normative status of the stricture that evidence evaluation not be contaminated by prior belief. Kornblith (1993; see also, Koehler, 1993) argued that

mistaken beliefs will, as a result of belief perseverance, taint our perception of new data. By the same token, however, belief perseverance will serve to color our perception of new data when our preexisting beliefs are accurate If, overall, our belief-generating mechanisms give us a fairly accurate picture of the world, then the phenomenon of belief perseverance may do more to inform our understanding than it does to distort it (p. 105).

Evans, Over, and Manktelow (1993) provided a similar account of belief bias in syllogistic reasoning. They considered the status of selective scrutiny explanations of the belief-bias phenomenon. Such theories posit that participants accept conclusions that are believable without engaging in logical reasoning at all. Only when faced with unbelievable conclusions do participants engage in logical reasoning about the premises. Evans et al. (1993) considered whether such a processing strategy could be rational in the sense of serving to achieve the person's goals, and they concluded that it could. They argued that any adult is likely to hold a large number of true beliefs that are interconnected in complex ways. Because single-belief revision has interactive effects on the rest of the belief network, it may be computationally costly (see Cherniak, 1986; Harman, 1986, 1995). Evans et al. (1993) argued that under such conditions it is quite right that conclusions that contradict one's beliefs "should be subjected to the closest possible scrutiny and refuted if at all possible" (p. 174).

Accounts of belief bias such as those of Kornblith (1993) and Evans et al. (1993) are in the tradition of optimization models (Schoemaker, 1991) that emphasize the adaptiveness of human cognition (Anderson, 1990, 1991; Campbell, 1987; Cosmides & Tooby, 1994, 1996; Friedrich, 1993; Oaksford & Chater, 1993, 1994, 1995). For example, Anderson (1990, 1991)—building on the work of Marr (1982), Newell (1982), and Pylyshyn (1984)—defined four levels of theorizing in cognitive science: a biological level that is inaccessible to cognitive theorizing, an implementation level designed to approximate the biological, an algorithmic level (an abstract specification of the steps necessary to carry out a process), and the rational level. The last level provides a specification of the goals of the system's computations (what the system is attempting to compute and why) and can be used to suggest constraints on the operation of the algorithmic level. According to Anderson (1990), the rational level specifies what are the "constraints on the behavior of the system in order for that behavior to be optimal" (p. 22). The description of this level of analysis proceeds from a "general principle of rationality" that assumes that "the cognitive system operates at all times to optimize the adaptation of the behavior of the organism" (Anderson, 1990, p. 28). Thus, the rational level of analysis concerns the goals of the system, the beliefs relevant to those goals, and the choice of action that is rational given the system's goals and beliefs (Anderson, 1990; Bratman, Israel, & Pollack, 1991; Dennett, 1987; Newell, 1982, 1990).

However, even if humans are optimally adapted to their environments at the rational level of analysis, there may still be computational limitations at the algorithmic level that prevent the full realization of the optimal model (e.g., Cherniak, 1986; Goldman, 1978; Oaksford & Chater, 1993, 1995). Assuming that the rational model for all humans in a given environment is the same, one would still expect individual differences in actual performance (despite no rational-level differences) due to differences at the algorithmic level. The responses of organisms with fewer algorithmic limitations would be assumed to be closer to the response that a rational analysis would reveal as optimal. Thus, the direction of the correlation between response type and cognitive capacity provides an empirical clue about the nature of the optimizing response. Specifically, if the standard position in the critical thinking literature is correct, then people with greater algorithmic capacity should show an enhanced ability to reason independently of prior belief. In contrast, if the critiques of the emphasis on decontextualization are correct, then this association should be attenuated.

Cognitive Capacity and Thinking Dispositions

The distinction between the algorithmic and rational levels of analysis made by computational theorists is similar to the distinction between cognitive capacity and thinking dispositions made by many psychological theorists (e.g., Baron, 1985, 1988, 1993; Ennis, 1987; Moshman, 1994;

Norris, 1992; Perkins et al., 1993). We shall use this distinction in our analysis of performance on an argument evaluation task in which belief bias must be avoided.

Cognitive capacities refer to the type of cognitive processes studied by information processing researchers who are seeking the underlying cognitive basis of performance on IQ tests. Perceptual speed, discrimination accuracy, working memory capacity, and the efficiency of the retrieval of information stored in long-term memory are examples of cognitive capacities that underlie traditional psychometric intelligence (Carpenter, Just, & Shell, 1990; Estes, 1982; Hunt, 1978, 1987; Lohman, 1989). As proxies for these processes, we used two global indicators of cognitive ability—Scholastic Aptitude Test (SAT) scores and a vocabulary test—that are known to correlate with a variety of cognitive subprocesses (Carroll, 1993; Matarazzo, 1972) that have been linked to neurophysiological and information processing indicators of efficient cognitive functioning (Caryl, 1994; Deary, 1995; Deary & Stough, 1996; Detterman, 1994; Hunt, 1987; Vernon, 1991, 1993; Vernon & Mori, 1992).

In contrast to cognitive capacities, thinking dispositions are best viewed as more malleable cognitive styles (see Baron, 1985). Because thinking dispositions and cognitive capacity are thought to differ in their degree of malleability, it is important to determine the relative proportion of variance in critical thinking skills (such as argument evaluation) that can be explained by each. To the extent that thinking dispositions explain variance in a critical thinking skill independent of cognitive capacity, theorists such as Baron (1985, 1988, 1993) would predict that the skill would be more teachable.

Further, such a partitioning is theoretically important because it is likely that the two constructs (cognitive ability and thinking dispositions) are at different levels of analysis in a cognitive theory. Variation in cognitive ability refers to individual differences in the efficiency of processing at the algorithmic level, whereas thinking dispositions of the type that we studied in this investigation elucidate individual differences at the rational level of analysis (Anderson, 1990; Stanovich, in press)—they index differences in goals and epistemic values. Many cognitive processes have been analyzed with concepts from the algorithmic level exclusively (Baron, 1985), and this has often impeded understanding. As educational psychologists become more concerned with thinking styles that are at the interface of cognitive and personality theory (see Goff & Ackerman, 1992; Keating, 1990; Moshman, 1994; Perkins et al., 1993; Sternberg & Ruzgis, 1994) it is critical that analyses of cognitive tasks encompass the rational as well as the algorithmic level of analysis.

A New Analytic Strategy for Assessing Argument Evaluation Ability

In this investigation, we employed the cognitive capacity–thinking disposition distinction as a framework for examining individual differences in the ability to reason about argument structure and avoid potential belief bias. We

also introduce an analytic technique for developing separate indices of a person's reliance on the quality of an argument and on their prior personal beliefs about the issue in question. Our methodology involved assessing, on a separate instrument, the participant's prior beliefs about a series of propositions. On an argument evaluation measure, administered at a later time, the participants evaluated the quality of arguments related to the same propositions. The arguments had an operationally determined objective quality that varied from item to item. Our analytic strategy was to regress each participant's evaluations of the argument simultaneously, on the objective measure of argument quality and on the strength of the belief that he or she had about the propositions prior to reading the argument. The standardized beta weight for argument quality then becomes an index of that participant's reliance on the quality of the arguments independent of their beliefs on the issues in question. Conversely, the standardized beta weight for the participant's prior belief reflects the biasing effects of beliefs on judgments of argument quality for that particular participant. The magnitude of the former statistic becomes an index of argument-driven, or data-driven processing (to use Norman's [1976] term), and the magnitude of the latter statistic becomes an index of the extent of belief-driven processing.

Our methodology is different from the traditional logic used in critical thinking tests, and it is a more sensitive one for measuring individual differences. For example, those who devise standardized critical thinking tests often simply try to balance opinions across items by using a variety of issues and by relying on chance to ensure that prior belief and strength of the argument are relatively balanced from respondent to respondent (Watson & Glaser, 1980). In contrast, we actually measured the prior opinion and took it into account in the analysis (for related techniques, see Klaczynski & Gordon, 1996; Kuhn, 1991, 1993; Kuhn, Amsel, & O'Loughlin, 1988; Slusher & Anderson, 1996). The technique allowed us to examine thought processes in areas of "hot" cognition in which biases are most likely to operate (Babad, 1987; Babad & Katz, 1991; Granberg & Brent, 1983; Kunda, 1990; Pyszczynski & Greenberg, 1987).

Research Plan

We first used the analytic logic described above on an experimental instrument we call the Argument Evaluation Test (AET). Subsequent to isolating individual differences in data-driven versus belief-driven thinking on the AET, we present a preliminary examination of the correlates of these individual differences. We examined whether the primary index of data-driven processing is correlated with measures of cognitive ability. As proxies for the latter, we used SAT scores and a vocabulary test.¹ We also examined whether

¹ As mentioned previously, tests such as these have been linked to neurophysiological and information processing indicators of efficient cognitive functioning (e.g., Deary, 1995; Hunt, 1987; Vernon, 1993).

thinking dispositions can account for variance in performance on the AET after cognitive ability has been partialled out (and the converse). In these analyses, we focused on whether we needed to implicate the rational level of explanation (corresponding to thinking dispositions) in addition to computational capacity to explain bias-free reasoning skill.

Within the larger domain of thinking dispositions, we examined those that we viewed as most relevant to rational thought, which we conceptualize as thought processes leading to more accurate belief formation, to more consistent belief-desire networks, and to more optimal actions (Audi, 1993a, 1993b; Baron, 1994; Gauthier, 1986; Gibbard, 1990; Goldman, 1986; Harman, 1995; Nozick, 1993; Stanovich, 1994, in press; Thagard, 1992). In our sampling of thinking dispositions to investigate, we focused in particular on thinking dispositions with potential epistemic significance, for example, "the disposition to weigh new evidence against a favored belief heavily (or lightly), the disposition to spend a great deal of time (or very little) on a problem before giving up, or the disposition to weigh heavily the opinions of others in forming one's own" (Baron, 1985, p. 15). Baron (1985, 1988, 1993) has called such tendencies dispositions toward actively open-minded thinking.

We constructed our thinking dispositions measure by drawing on scales already described in the literature and by constructing some of our own. Several of our measures had strong similarities to other related measures in the literature (Cacioppo, Petty, Feinstein, & Jarvis, 1996; Schommer, 1990, 1993, 1994; Schommer & Walker, 1995; Webster & Kruglanski, 1994). Perhaps the strongest similarities are with the two dispositional factors that Schommer (1990, 1993) called "belief in simple knowledge" and "belief in certain knowledge." However, aspects of the need for cognition (Cacioppo et al., 1996) and the need for closure (Kruglanski & Webster, 1996) constructs are also represented in our scale. Overall, we attempted to tap the following dimensions: epistemological absolutism, willingness to perspective-switch, willingness to decontextualize, and the tendency to consider alternative opinions and evidence.

Method

Participants

The participants were 349 undergraduate students (134 men and 215 women) recruited through an introductory psychology participant pool at a medium-sized state university. Their mean age was 18 years 10 months ($SD = 2.2$). There were no gender differences on any of the experimental measures, so this variable will not be considered further.

Argument Evaluation Test

The argument evaluation test (AET) consisted of two parts. In the first, participants indicated their degree of agreement with a series of 23 target propositions on a 4-point scale: 4 (*strongly agree*), 3 (*agree*), 2 (*disagree*) and 1 (*strongly disagree*). With one exception, the propositions all concerned real social and political

issues on which people hold varying, and sometimes strong, beliefs (e.g., gun control, taxes, university governance, crime, automobile speed limits). For example, the following target proposition was in one item: "The welfare system should be drastically cut back in size." Scores on this part of the questionnaire sometimes are referred to in an abbreviated fashion as prior belief scores. The AET prior belief items varied greatly in the degree to which the sample as a whole endorsed them—from a low of 1.79 for the item, "It is more dangerous to travel by air than by car" to a high of 3.64 for the item, "Seat belts should always be worn when traveling in a car."

After completing several other questionnaires and tasks, the participants completed the second part of the AET. The instructions introduced the participants to a fictitious individual, Dale, whose arguments they were to evaluate. Each of the 23 items on the second part of the instrument began with Dale stating a belief about an issue. The 23 beliefs were identical to the target propositions that the participants had rated their degree of agreement with on the prior belief part of the instrument (e.g., "The welfare system should be drastically cut back in size"). Dale then provided a justification for the belief (in this case, e.g., "The welfare system should be drastically reduced in size because welfare recipients take advantage of the system and buy expensive foods with their food stamps"). A critic then presents an argument to counter this justification (e.g., "Ninety-five percent of welfare recipients use their food stamps to obtain the bare essentials for their families"). The participant is told to assume that the facts in the counterargument are correct. Finally, Dale attempts to rebut the counterargument (e.g., "Many people who are on welfare are lazy and don't want to work for a living"). Again assuming that the facts in the rebuttal are correct, the participant is told to evaluate the strength of Dale's rebuttal to the critic's argument. The instructions remind the participant that he or she is to focus on the quality of Dale's rebuttal and to ignore whether or not he or she agreed with Dale's original belief. The participant then evaluated the strength of the rebuttal on a 4-point scale: 4 (*very strong*), 3 (*strong*), 2 (*weak*), and 1 (*very weak*). The remaining 22 items were structured analogously. A deliberate attempt was made to vary the quality of the rebuttals. The mean evaluation of the rebuttals ranged from 1.93 to 3.41 across the 23 items.

The instructions and several examples of items from the second part of the AET are presented in Appendix A. The full set of AET items, as well as the items on the Thinking Dispositions Questionnaire (to be described below) are available from Keith E. Stanovich on request.

The analysis of performance on the AET required that the participants' evaluations of argument quality be compared to some objective standard. We used a summary measure of eight experts' evaluations of these rebuttals as an operationally defined standard of argument quality. Specifically, three full-time faculty members of the Department of Philosophy at the University of Washington, three full-time faculty members of the Department of Philosophy at the University of California, Berkeley, and the two authors judged the strength of the rebuttals. The median correlation between the judgments of the eight experts was .74. Although we devised the problems, the median correlations between our judgments and those of the external experts were reasonably high (.78 and .73, respectively) and roughly equal to the median correlation among the judgments of the six external experts themselves (.73). Thus, for the purposes of the regression analyses described below, the median of the eight experts' judgments of the rebuttal quality served as the objective index of argument quality for each item. These scores will sometimes be referred to in an abbreviated fashion as the argument quality variable. This variable ranged from

1.0 to 4.0 across the 23 items. Eight rebuttals received a score of 1.0, three received a score of 1.5, three received a score of 2.0, two received a score of 2.5, four received a score of 3.0, and three received a score of 4.0.

As an example, on the above item, the median of the experts' ratings of the rebuttal was 1.5 (between *weak* and *very weak*). The mean rating that the participants gave the item was 1.93 (*weak*), although the participants' mean prior belief score indicated a neutral opinion (2.64). Our attempt to construct items so that objectively strong and weak arguments were equally associated with the prior beliefs that we thought participants would endorse was very successful, as there was a nonsignificant .01 correlation between the sample's mean prior belief scores and objective argument quality. Further details of the analysis of this task are presented in the Results section.

One indication of the validity of the experts' ratings is that the experts were vastly more consistent among themselves in their evaluation of the items than were the participants. Because the median correlation among the eight experts' judgments was .74, a parallel analysis of consistency among the participants was conducted to provide a comparison. Forty-three groups of 8 participants were formed, and the correlations among the argument evaluations for the 8 individuals in each group were calculated. The median correlation for each of the 43 groups was then determined. The highest median across all of the 43 groups was .59. Thus, not one of the 43 groups attained the level of agreement achieved by the eight experts. The mean of the median correlations in the 43 groups was .28, markedly below the degree of consistency in the experts' judgments.

General Ability Measures

SAT scores. Because SAT scores were not available to us because of university restrictions, students were asked to indicate their verbal and mathematical SAT scores on a demographics sheet. The mean reported verbal SAT score of the students was 529 ($SD = 72$), the mean reported mathematical SAT score was 578 ($SD = 72$), and the mean total SAT score was 1107 ($SD = 108$). These reported scores match the averages of this institution (520, 587, and 1107) quite closely (Straughn & Straughn, 1995; all SAT scores were from administrations of the test prior to its recent rescaling). A further indication of the validity of the self-reported scores is that the correlation between a vocabulary test (described below) and the reported SAT total score (.49) was quite similar to the .51 correlation between the vocabulary checklist and verified total SAT scores in a previous investigation with the same vocabulary measure (West & Stanovich, 1991). A final indication of the validity of the SAT reports is that the vocabulary test displayed a higher correlation with the verbal SAT scores (.61) than with the mathematical SAT scores (.13). The difference between these dependent correlations (see Cohen & Cohen, 1983, pp. 56–57) was highly significant ($p < .001$).

Vocabulary test. As an additional converging measure of cognitive ability to supplement the SAT scores, a brief vocabulary measure was administered to the participants (because vocabulary is the strongest specific correlate of general intelligence, see Matarazzo, 1972). This task employed the checklist-with-foils format that has been shown to be reliable and valid in assessing individual differences in vocabulary knowledge (Anderson & Freebody, 1983; Cooksey & Freebody, 1987; White, Slater, & Graves, 1989; Zimmerman, Broder, Shaughnessy, & Underwood, 1977). The stimuli for the task were 40 words and 20 pronounceable nonwords taken largely from the stimulus list of Zimmerman et al. (1977). The words and nonwords were intermixed through alphabetization.

The participants were told that some of the letter strings were actual words and that others were not and that their task was to read through the list of items and to put a check mark next to those that they knew were words. Scoring on the task was determined by taking the proportion of the target items that were checked and subtracting the proportion of foils checked. Other corrections for guessing and differential criterion effects (see Snodgrass & Corwin, 1988) produced virtually identical correlational results. The scores ranged from .100 to .925. The split-half reliability of the number of correct items checked was .87 (Spearman-Brown corrected).

Thinking Dispositions Questionnaire

Participants completed a questionnaire consisting of a number of subscales. The response format for each item in the questionnaire (with the exception of the outcome bias items) was 6 (*agree strongly*), 5 (*agree moderately*), 4 (*agree slightly*), 3 (*disagree slightly*), 2 (*disagree moderately*), and 1 (*disagree strongly*). The items from the subscales were randomly intermixed in the questionnaire. The subscales were as follows:

Flexible Thinking scale. We devised the items on the Flexible Thinking scale. The design of the items was influenced by a variety of sources from the critical thinking literature (e.g., Ennis, 1987; Facione, 1992; Nickerson, 1987; Norris & Ennis, 1989; Perkins et al., 1993; Zechmeister & Johnson, 1992) but most specifically by the work of Baron (1985, 1988, 1993). Baron has emphasized that actively open-minded thinking is a multifaceted construct encompassing the cultivation of reflectiveness rather than impulsivity, the seeking and processing of information that disconfirms one's belief (as opposed to confirmation bias in evidence seeking), and the willingness to change one's beliefs in the face of contradictory evidence. There were 10 items on the Flexible Thinking scale, some tapping the disposition toward reflectivity ("If I think longer about a problem I will be more likely to solve it;" "Difficulties can usually be overcome by thinking about the problem, rather than through waiting for good fortune;" "Intuition is the best guide in making decisions;" and "Coming to decisions quickly is a sign of wisdom"—the latter two reverse scored), willingness to consider evidence contradictory to beliefs (e.g., "People should always take into consideration evidence that goes against their beliefs"), willingness to consider alternative opinions and explanations ("A person should always consider new possibilities," "Considering too many different opinions often leads to bad decisions," the latter reverse scored), and a tolerance for ambiguity combined with a willingness to postpone closure ("There is nothing wrong with being undecided about many issues," "Changing your mind is a sign of weakness," and "Basically, I know everything I need to know about the important things in life,"—the latter two reverse scored). The scores on this scale ranged from 30 to 59. The split-half reliability of the scale was .49 (Spearman-Brown corrected), and Cronbach's alpha was .50.

Openness-Ideas subscale. The eight items from the Openness-Ideas facet of the Revised NEO Personality Inventory (Costa & McCrae, 1992) were administered (e.g., "I have a lot of intellectual curiosity;" "I find philosophical arguments boring"—the latter reverse scored). The scores on this scale ranged from 15 to 48. The split-half reliability of the scale was .73 (Spearman-Brown corrected), and Cronbach's alpha was .77.

Openness-Values subscale. The eight items from the Openness-Values facet of the Revised NEO Personality Inventory were administered (e.g., "I believe that laws and social policies should change to reflect the needs of a changing world;" "I believe letting students hear controversial speakers can only confuse and

mislead them"—the latter reverse scored). The scores on this scale ranged from 13 to 48. The split-half reliability of the scale was .73 (Spearman-Brown corrected), and Cronbach's alpha was .71.

Absolutism subscale. This scale was adapted from the Scale of Intellectual Development (SID) developed by Erwin (1981, 1983). The SID represents an attempt to develop a multiple-choice scale to measure Perry's (1970) hypothesized stages of epistemological development in young adulthood. It is similar to related work in the literature (King & Kitchener, 1994; Kramer, Kahlbaugh, & Goldston, 1992; Schommer, 1990, 1993, 1994). We chose nine items designed to tap Perry's early stages that are characterized by an absolutist orientation. These early stages are characterized by cognitive rigidity, by a belief that issues can be couched in either-or terms, by a belief that there is one right answer to every complex problem, and by reliance on authority for belief justification. This orientation is captured by items such as "In most situations requiring a decision, it is better to listen to someone who knows what they are doing;" "It is better to simply believe in a religion than to be confused by doubts about it;" "Right and wrong never change;" and "I can't enjoy the company of people who don't share my moral values." The scores on this scale ranged from 9 to 46. The split-half reliability of the scale was .69 (Spearman-Brown corrected), and Cronbach's alpha was .64.

Dogmatism subscale. The Dogmatism subscale consisted of nine items. Three were taken from a short-form field version (Troidahl & Powell, 1965) of Rokeach's (1960) Dogmatism scale: "Of all the different philosophies which exist in the world there is probably only one which is correct;" "Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups;" and "There are two kinds of people in this world: those who are for the truth and those who are against the truth". Two items were drawn from the Dogmatic Thinking scale used by Paulhus and Reid (1991): "Often, when people criticize me, they don't have their facts straight" and "No one can talk me out of something I know is right." Four items were taken from the full Rokeach scale published in Robinson, Shaver, and Wrightsman (1991): "A group which tolerates too much difference of opinion among its members cannot exist for long;" "There are a number of people I have come to hate because of the things they stand for;" "My blood boils over whenever a person stubbornly refuses to admit he's wrong;" and "Most people just don't know what's good for them." The scores on this scale ranged from 11 to 44. The split-half reliability of the scale was .54 (Spearman-Brown corrected), and Cronbach's alpha was .60.

Categorical Thinking subscale. Three items from the Categorical Thinking subscale of Epstein and Meier's (1989) Constructive Thinking Inventory were administered: "There are basically two kinds of people in this world, good and bad;" "I think there are many wrong ways, but only one right way, to almost anything;" and "I tend to classify people as either for me or against me." The scores on this scale ranged from 3 to 16.

Superstitious Thinking subscale. This task was included because of our longstanding interest in the correlates of paranormal belief (Stanovich, 1989). The seven items on this scale consisted of four items concerning the concept of luck (e.g., "I have personal possessions that bring me luck at times;" "The number 13 is unlucky") adapted from the Paranormal Beliefs Questionnaire developed by Tobacyk and Milford (1983) and three items from the Superstitious Thinking subscale of Epstein and Meier's (1989) Constructive Thinking Inventory. The scores on this scale ranged from 7 to 38. The split-half reliability of the scale was .73 (Spearman-Brown corrected), and Cronbach's alpha was .73.

Counterfactual Thinking scale. As an indicator of the ability to

decenter and adopt alternative perspectives, we devised a two-item scale designed to tap counterfactual thinking. The two scale items were "My beliefs would not have been very different if I had been raised by a different set of parents;" and "Even if my environment (family, neighborhood, schools) had been different, I probably would have the same religious views." Both items were reversed scored so that higher scores indicated counterfactual thinking. The scores on this scale ranged from 2 to 12.

Outcome Bias subscale. As an indicator of the disposition toward decontextualization, we chose the phenomenon of outcome bias. Our measure of outcome bias derived from a problem investigated by Baron and Hershey (1988):

A 55-year-old man had a heart condition. He had to stop working because of chest pain. He enjoyed his work and did not want to stop. His pain also interfered with other things, such as travel and recreation. A successful bypass operation would relieve his pain and increase his life expectancy from age 65 to age 70. However, 8% of the people who have this operation die from the operation itself. His physician decided to go ahead with the operation. The operation succeeded. Evaluate the physician's decision to go ahead with the operation.

Participants responded on a seven-point scale ranging from *incorrect, a very bad decision* to *clearly correct, an excellent decision*. This question appeared early in the questionnaire. Later in the questionnaire, participants evaluated a different decision to perform surgery that was designed to be objectively better than the first (2% chance of death rather than 8%, 10-year increase in life expectancy versus 5-year increase, etc.) even though it had an unfortunate negative outcome (death of the patient). Participants reported on the same scale as before. An outcome bias was demonstrated when these two items were compared: 118 participants rated the positive outcome decision as better than the negative outcome decision, 178 participants rated the two decisions equally, and 53 participants rated the negative outcome decision as the better decision. The measure of outcome bias that we used was the scale value of the decision on the positive outcome case (from 1 to 7) minus the scale value of the negative outcome decision. The higher the value, the more the outcome bias. The mean outcome bias score was .358, which was significantly different from zero, $t(348) = 5.43, p < .001$. We interpret low outcome bias scores as indicating greater decontextualization.

Social Desirability Response Bias subscale. Five items reflecting social desirability response bias (Furnham, 1986; Paulhus & Reid, 1991) were taken from the Balanced Inventory of Desirable Responding (Paulhus, 1991): "I always obey laws, even if I'm unlikely to get caught;" "I don't gossip about other people's business;" "I sometimes tell lies if I have to;" "There have been occasions when I have taken advantage of someone;" and "I have said something bad about a friend behind his or her back." The scores on this scale ranged from 5 to 26.

Composite Actively Open-Minded Thinking score. Because several of the scales displayed moderate intercorrelations (see Appendix B), a composite actively open-minded thinking (AOT) score was formed by summing the scores on the Flexible Thinking, Openness-Ideas, and Openness-Values scales and subtracting the sum of the Absolutism, Dogmatism, and Categorical Thinking scales. Thus, high scores on the AOT composite indicate openness to belief change and cognitive flexibility, whereas low scores indicate cognitive rigidity and resistance to belief change. The creation of this composite scale was justified by a principal components analysis conducted on the eight subscales (these six subscales plus the Superstitious Thinking and Counterfactual Think-

ing scales). Two components had eigenvalues greater than one. The loadings, displayed in Appendix C, are subsequent to varimax rotation. The first component, which accounted for 40.8% of the variance, received high loadings from each of the six subscales in the AOT score. Counterfactual thinking, which failed to correlate with the other variables, primarily defined the second component that accounted for only 13.5% of the variance and barely made the eigenvalue >1 criterion (1.083). Use of a factor score for the first component produced results virtually identical to those from the unit-weighted composite; therefore we have used the latter, more parsimonious index.² The scores on this scale ranged from -9 to 112. The split-half reliability of the composite scale was .90 (Spearman-Brown corrected), and Cronbach's alpha was .88.

Procedure

Participants completed the tasks during a single 2-hour session in which they also completed some other tasks that were not part of this investigation. They were tested in small groups of 3 to 8 individuals. The order of administration of the tasks subsequent to filling out the demographics sheet was outcome bias (Part 1), the AET prior belief section, Thinking Dispositions Questionnaire, outcome bias (Part 2), vocabulary test, and the AET evaluation section.

Results

Individual differences in participants' relative reliance on objective argument quality and prior belief were examined by running separate regression analyses on each participant's responses. That is, a separate multiple regression equation was constructed for each participant. The participant's evaluations of argument quality served as the criterion variable in each of 349 separate regression analyses. The 23 evaluation scores were regressed simultaneously on both the 23 argument quality scores and the 23 prior belief scores. Thus, a total of 349 individual regression analyses (one for each participant) were conducted. For each participant, these analyses resulted in two beta weights—one for argument quality and one for prior belief. The former beta weight—an indication of the degree of reliance on argument quality independent of prior belief—is the primary indicator of the ability to evaluate arguments independent of one's beliefs.

The mean multiple correlation across the 349 separate regressions in which each participant's evaluations were regressed on the objective argument quality scores and on his or her prior belief scores was .451 ($SD = .158$). Across the 349 regressions, the mean standardized beta weight for argument quality was .330 ($SD = .222$). These latter values varied widely—from a low of $-.489$ to a high of $.751$. However, the variation was asymmetric around zero, as is clear from Figure 1, which plots the frequency distribution of the standardized beta weights for argument quality. Only 30 of 349 participants had beta weights less than zero.

Across the 349 regressions, the mean standardized beta weight for the participant's prior belief scores was .151 ($SD = .218$). Although low, this mean was significantly different from zero, $t(348) = 12.93, p < .001$. Nevertheless, these values also varied widely—from a low of $-.406$ to a

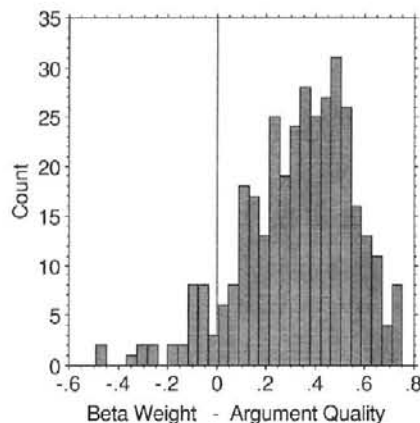


Figure 1. Frequency distribution of the beta weights for argument quality in the multiple regressions conducted on each participant's responses on the Argument Evaluation Test.

high of $.662$, as indicated in Figure 2, which displays a plot of the frequency distribution of the standardized beta weights. Thus, as both figures indicate, individuals vary substantially in their reliance on argument quality and prior belief when evaluating the arguments.

As mentioned previously, we view the beta weight for argument quality as the primary measure of the participant's ability to reason independently of their own beliefs. Thus, we explored the correlates of this ability by splitting the sample in half on the basis of this index.³ A median split of the sample resulted in a subsample of 174 participants with mean beta weights for argument quality of $.504$ (termed HIARG because of their high reliance on argument quality) and 175 low scoring participants having mean argument quality beta weights of $.156$ (termed LOARG because of their low reliance on argument quality). The difference in beta weights was highly significant, $t(347) = 23.56, p < .001$. The mean beta weight for prior belief in the HIARG group ($.128$) was also significantly lower than the beta weight for prior belief in the LOARG group ($.173$), $t(347) = 1.98, p < .05$, indicating that this median split reliably partitioned the sample into a group of participants relatively more reliant on argument quality for their argument evaluation decisions (HIARG) and a group relatively more reliant on their prior belief about the issue when making argument evaluation decisions (LOARG).

As indicated in Table 1, the mean SAT total scores of the HIARG group were significantly ($p < .001$) higher than

² The creation of this composite scale was also justified by a principal components analysis conducted on the 56 items from all of the scales. The dominant first principal component was primarily composed of high loadings for the items from the six scales comprising the AOT score. Other items displayed small loadings on this component and instead contributed to several of the minor components.

³ We obtained virtually identical results by using as a variable the beta weight for argument quality minus the beta weight for prior belief. Therefore, the simpler index—the beta weight for argument quality—was employed in the subsequent analyses.

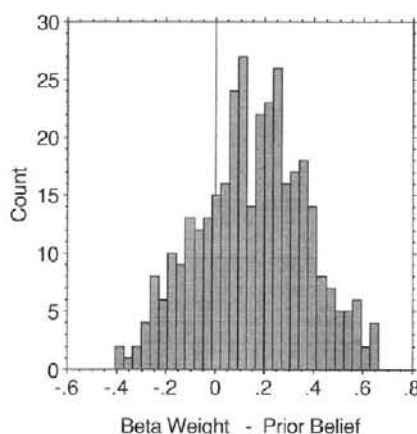


Figure 2. Frequency distribution of the beta weights for prior belief in the multiple regressions conducted on each participant's responses on the Argument Evaluation Test.

those of the LOARG group, indicating that the group more reliant on argument quality for their judgments was higher in cognitive ability. This finding was confirmed by the existence of differences in favor of the HIARG group on the vocabulary test ($p < .001$).

The remainder of Table 1 presents comparisons of the two groups on the different scales of the Thinking Dispositions Questionnaire. The HIARG group scored significantly higher on the Flexible Thinking scale, Openmindedness-Ideas, and Openmindedness-Values scales and significantly lower on the Absolutism, Dogmatism, and Categorical Thinking scales. The AOT composite variable, formed by a linear combination of these six scales (forming the composite from standardized scores of the variables resulted in virtually identical results), displayed a highly significant ($p < .001$) difference. The HIARG group also displayed significantly lower scores on the Superstitious Thinking scale, significantly higher scores on the two-item Counterfactual Thinking scale, and showed significantly less outcome bias. None of the response patterns were due to a tendency for the HIARG group to give more socially desirable responses because the groups did not differ on that variable.

As the results displayed in Table 1 indicate, a consistent pattern of differences on the thinking dispositions measures was observed. The group of participants who displayed more reliance on argument quality in their judgments on the Argument Evaluation Test consistently displayed more openmindedness, cognitive flexibility, and skepticism, and less dogmatism and cognitive rigidity. However, as Table 1 also indicates, the groups differed in cognitive ability (SAT and vocabulary scores). Thus, it is possible that the differences in thinking dispositions are merely a function of cognitive ability differences. The analyses displayed in Table 2 were designed to address the question of whether differences in thinking dispositions between the groups were entirely attributable to differences in cognitive ability. These analyses were structured to examine the question of whether thinking dispositions can predict performance on

the AET even after cognitive ability has been partialled out.

The criterion variable in the series of hierarchical regression analyses presented in Table 2 was the beta weight for argument quality on the AET. Entered first into the regression equation was the SAT total score which, not surprisingly, given the results displayed in Table 1, accounted for a significant proportion of variance.⁴ However, listed next are alternative second steps in the hierarchical analysis. As the table indicates, the AOT composite score accounted for significant variance in performance on the AET even after SAT scores were entered into the equation—as did scores on the superstitious thinking, outcome bias (a measure of decontextualization), and counterfactual thinking (a measure of perspective switching) measures. Thus, the linkage between thinking dispositions and performance on the AET illustrated in Table 1 is not entirely due to covariance with cognitive ability. Various measures of thinking dispositions are predictors, independent of SAT scores.

The second analysis displayed in Table 2 forces in AOT scores as a predictor first. Its zero-order correlation with the beta weight for argument quality (.292) is somewhat lower than that for SAT scores (.353) but is of the same order of magnitude. The table indicates that after AOT scores are in the equation, the measure of outcome bias accounted for significant variance when entered as the second step. The measure of counterfactual reasoning accounted for significant additional variance when entered as the third step, but the superstitious thinking measure did not account for significant additional variance when entered as the fourth—however, the multiple R of the dispositions measures, .341, now approaches that for SAT scores. The latter, when entered as the fifth step, accounted for 7.8% unique variance ($p < .001$) indicating that this measure of cognitive ability accounts for substantial additional variance after all of the variance associated with the thinking dispositions measures has been partialled out.

A stepwise regression analysis with these same variables was also conducted. The first variable to enter was SAT scores, which accounted for 12.4% of the variance ($p < .001$). The next variable to enter the equation was the AOT composite, which accounted for 3.7% additional variance ($p < .001$). Entering third was the outcome bias variable, which accounted for 1.7% additional variance ($p < .05$). The last variable to explain significant unique variance was the measure of counterfactual thinking, which explained 1.4% additional variance ($p < .05$). (The beta weights of all

⁴ A further validation of the experts' ratings comes from comparing the zero-order correlation of SAT scores and the beta weight for the experts' ratings (.35) with the zero-order correlation of SAT scores and the beta weights based on the mean ratings of argument quality given by the participants (.23). This difference in dependent correlations (see Cohen & Cohen, 1983, pp. 56–57) was statistically significant, $t(346) = 3.20$, $p < .01$, indicating that participants of higher cognitive ability were more likely to agree with the experts in their judgments of argument quality than with their fellow participants of lower cognitive ability.

Table 1
Mean Scores of Participants With Highest ($n = 174$) and Lowest ($n = 175$) Argument Quality Scores on the Argument Evaluation Test (AET; Standard Deviations in Parentheses)

Variable	LOARG	HIARG	<i>t</i>
AET-argument quality (β weight)	.156 (.17)	.504 (.10)	23.56***
AET-prior belief (β weight)	.173 (.24)	.128 (.19)	1.98*
SAT total	1080 (106)	1133 (103)	4.64***
Vocabulary checklist	.528 (.160)	.592 (.150)	3.89***
Thinking dispositions			
Flexible Thinking scale	44.2 (4.6)	45.5 (4.9)	2.59*
Openness-Ideas	32.5 (6.3)	34.6 (6.4)	3.00**
Openness-Values	34.4 (5.4)	36.8 (5.4)	4.06***
Absolutism scale	30.1 (5.7)	28.0 (6.1)	-3.25**
Dogmatism scale	30.5 (5.5)	28.2 (5.7)	-3.84***
Categorical Thinking	8.2 (2.7)	7.0 (2.6)	-4.05***
AOT composite	42.3 (21.7)	53.6 (22.9)	4.71***
Superstitious Thinking	18.2 (6.0)	16.7 (5.4)	-2.36*
Counterfactual Thinking	8.2 (2.6)	9.0 (2.4)	3.12**
Outcome Bias	.583 (1.3)	.144 (1.2)	-3.33***
Social Desirability scale	14.9 (3.6)	15.2 (3.8)	0.84

Note. $df = 347$ for the AET and vocabulary checklist, 338 for the SAT total, and 346 for the Thinking Dispositions Scales. LOARG = low reliance on argument quality; HIARG = high reliance on argument quality; SAT = Scholastic Aptitude Test; AOT composite = (Flexible Thinking scale + Openness-Ideas + Openness-Values) - (Absolutism scale + Dogmatism scale + Categorical Thinking).

* $p < .05$. ** $p < .01$. *** $p < .001$, all two-tailed.

of these variables remained significant in the final simultaneous equation.)

An additional stepwise analysis was conducted with, as potential predictors, all of the component subscales of the AOT composite in addition to the variables in the previous analysis. The results were highly convergent. In this stepwise regression, the first variable to be entered was SAT scores, which accounted for 12.4% of the variance ($p < .001$). The next variable to be entered to the equation was

the Categorical Thinking subscale, which accounted for 3.7% additional variance ($p < .001$). Entering third was the outcome bias variable, which accounted for 1.8% additional variance ($p < .05$). The last variable to explain significant unique variance was the measure of counterfactual thinking, which explained 1.4% additional variance ($p < .05$). (The beta weights of all of these variables remained significant in the final simultaneous equation.)

We carried out a further, converging examination of the covariance relationships among performance on the AET, cognitive ability, and thinking dispositions by conducting a commonality analysis (Kerlinger & Pedhazur, 1973). The criterion variable in the analysis displayed in Table 3 was again the beta weight for argument quality. In the hierarchical analyses that estimate the commonality components (see Kerlinger & Pedhazur, 1973), cognitive ability and thinking dispositions were defined by sets of variables (see Cohen & Cohen, 1983). In this analysis, a more comprehensive index of cognitive ability was employed—the collective variance explained by the SAT and vocabulary scores. *Thinking dispositions* were defined as the collective variance explained by the AOT score, superstitious thinking, outcome bias, and counterfactual thinking measures. In total, 20.0% of the variance in AET performance was explained (multiple $R = .447$). A portion (8.4%) of this variance explained was unique to cognitive ability, and 6.7% of the variance was uniquely explained by the thinking dispositions. Only 4.9% of the variance explained was common to the cognitive ability and thinking dispositions measures. Thus, the commonality analysis indicates that there was considerable nonoverlap in the variance

Table 2
Hierarchical Regression Analyses Predicting Beta Weight for Argument Quality on the Argument Evaluation Test

Step	Variable	<i>R</i>	ΔR^2	ΔF
Analysis 1				
1	SAT total	.353	.124	47.86***
2	AOT composite	.401	.037	14.76***
2	Superstitious Thinking	.368	.012	4.36*
2	Outcome Bias	.391	.029	11.65***
2	Counterfactual Thinking	.380	.020	7.80**
Analysis 2				
1	AOT composite	.292	.085	31.42***
2	Outcome Bias	.315	.014	5.10*
3	Counterfactual Thinking	.334	.012	4.66**
4	Superstitious Thinking	.341	.005	1.89
5	SAT total	.441	.078	32.30***

* $p < .05$. ** $p < .01$. *** $p < .001$.

Note. SAT = Scholastic Aptitude Test; AOT composite = (Flexible Thinking scale + Openness-Ideas + Openness-Values) - (Absolutism scale + Dogmatism scale + Categorical Thinking).

Table 3
*Commonality Analysis Conducted With Beta Weight for
 Argument Quality on the Argument Evaluation Test as the
 Dependent Variable*

Variable	Variance		Total Variance Explained
	Unique	Common	
Thinking dispositions	.067	.049	.200
Cognitive ability	.084	.049	

Note. Thinking dispositions = variance accounted for by actively openminded thinking composite, superstitious thinking, outcome bias, and counterfactual thinking scores; Cognitive ability = variance accounted for by SAT and vocabulary scores.

in AET performance explained by cognitive ability and thinking dispositions. The unique variance attributable to each was larger than the common variance that they shared.

Discussion

The analytic logic that we used in designing the Argument Evaluation Test was intended to provide a means of separating prior belief from the assessment of how well an individual can track argument quality. It is our contention that the beta weight for argument quality in our analyses captures a quintessential aspect of critical thought: the ability to evaluate the quality of an argument independent of one's feelings and personal biases about the proposition at issue. We have shown that there is considerable variability in this index and that this variability has an orderly distribution (see Figure 1). Further, this variance can be reliably linked to individual differences in cognitive ability and to individual differences in thinking dispositions that have epistemic significance.

Of course, numerous caveats are in order. Our regression analyses estimated the beta weights from scores on only 23 items. A larger item set would presumably result in more stable estimates, and this result would be highly desirable. Nevertheless, the estimates were reliable enough to display significant associations with measures of cognitive ability and thinking dispositions. Also, regarding the latter, some of our scales were based on only a small number of items. Work with more extensive and refined scales is clearly needed, although again, numerous significant associations were detected with our new instruments and short scales.

Thinking Dispositions and Argument Evaluation

An important pattern in the analysis of individual differences was the relative separability of cognitive ability and thinking dispositions as predictors of argument evaluation performance. The regression and commonality analyses indicated that both were significant unique predictors and that their unique variance as predictors exceeded the variance that they had in common with each other. This finding supports conceptualizations of human cognition that empha-

size the potential separability of cognitive capacities and thinking dispositions as predictors of reasoning skill (e.g., Baron, 1985, 1988, 1993; Ennis, 1987; King & Kitchener, 1994; Kitchener & Fischer, 1990; Klaczynski, 1997; Norris, 1992; Siegel, 1993).

There are increasing indications that this separability is a pervasive and replicable phenomenon. Schommer (1990) found that a measure of belief in certain knowledge predicted the tendency to draw one-sided conclusions from ambiguous evidence even after verbal ability was controlled. Kardash and Scholes (1996) found that the tendency to properly draw inconclusive inferences from mixed evidence was related to belief in certain knowledge and need for cognition scores. Further, these relationships were not mediated by verbal ability because a vocabulary measure was essentially unrelated to evidence evaluation. Likewise, Klaczynski (1997; Klaczynski & Gordon, 1996) found that the degree to which adolescents criticized belief-inconsistent evidence more than belief-consistent evidence was unrelated to cognitive ability.

We found stronger relationships between cognitive ability and the tendency to evaluate evidence independent of prior beliefs than did these other studies, but there were numerous important task differences that could have accounted for this difference in results. For example, both Schommer (1990) and Kardash and Scholes (1996) presented their participants with conflicting evidence and arguments. Good reasoning was thus defined as coming to an uncertain or tentative conclusion. In our study, participants were not presented with conflicting arguments but instead with a single argument to be evaluated as either strong or weak. It might be that the cognitive processes involved in the two situations are somewhat different.

The study by Klaczynski and Gordon (1996) presents a different contrast. They presented participants with flawed hypothetical experiments that led to either belief-consistent or belief-inconsistent conclusions, and they then evaluated the quality of the reasoning participants used when they critiqued the flaws in the experiments. Klaczynski and Gordon (1996) found that verbal ability was related to the quality of the reasoning in both the belief-consistent and belief-inconsistent conditions. In their experiment, Klaczynski and Gordon observed a belief-bias effect—participants found more flaws when the experiment's conclusions were inconsistent than when they were consistent. However, verbal ability was not correlated with the magnitude of the belief-bias effect—even though it was correlated with overall levels of reasoning in each of the different conditions that were considered separately. It should be noted that the belief-bias difference score in the Klaczynski and Gordon study is not analogous to the regression coefficient for argument quality in the present study. A high score on the belief-bias difference score in the Klaczynski and Gordon study is a direct indicator of the magnitude of the belief-bias effect, whereas a low score for argument quality in our paradigm may come about in a variety of ways. High correlations of participant evaluations with prior beliefs do serve to reduce the beta weight for argument quality in our analyses. However, the beta weight for argument quality in

our paradigm will also be low if a participant is simply a poor argument evaluator.

In short, Klaczynski and Gordon's (1996) index is a direct measure of belief bias; ours is a measure of the ability to reason in situations in which prior beliefs may be interfering (i.e., reasoning in the face of potentially interfering prior beliefs). Our measure combines context-free reasoning ability with the ability to ignore belief bias. It is thus a more complex index (and clearly less desirable than a direct belief-bias index for some purposes). The fact that it is in part determined by context-free reasoning ability probably explains why we observed relationships with cognitive ability.

Although cognitive ability was implicated in performance on our task more strongly than it was in some other investigations (Kardash & Scholes, 1996; Schommer, 1990), our results converge with others strongly in indicating that epistemic dispositions can predict argument evaluation skill even when cognitive ability is partialled out. In the stepwise regression analyses, in addition to SAT scores, a measure of decontextualization (outcome bias), a measure of perspective switching (counterfactual thinking), and measures of epistemological absolutism (AOT composite and Categorical Thinking subscale) proved to be significant unique predictors.

Why might thinking dispositions continue to be unique predictors of AET performance even after cognitive ability has been partialled out? We posited earlier in this article that these two constructs (cognitive ability and thinking dispositions) are actually at different levels of analysis in a cognitive theory and that they do separate explanatory work. Variation in cognitive ability refers to individual differences in the efficiency of processing at the algorithmic level. In contrast, thinking dispositions of the type studied in this investigation elucidate individual differences at the rational level. These dispositions are telling us about the individual's goals and epistemic values (King & Kitchener, 1994; Kitcher, 1993; Kruglanski, 1989; Kruglanski & Webster, 1996; Pintrich, Marx, & Boyle, 1993; Schommer, 1990, 1993, 1994). For example, consider an individual who scores high on measures such as the Flexible Thinking scale and low on measures such as the Dogmatism and Absolutism scales—a person who agrees with statements such as "People should always take into consideration evidence that goes against their beliefs" and who disagrees with statements such as "No one can talk me out of something I know is right." Such a response pattern is indicating that changing beliefs to get closer to the truth is an important goal for this person. This individual is signaling that they value being an accurate belief forming system more than they value holding on to the beliefs that they currently have (see Cederblom, 1989 for an insightful discussion of this distinction). In contrast, consider a person scoring low on the Flexible Thinking scale and high on measures such as the Absolutism and Categorical Thinking scales—a person who disagrees with statements such as "A person should always consider new possibilities" and who agrees with statements such as "There are a number of people I have come to hate because of the things they stand for." Such a response

pattern is indicating that retaining current beliefs is an important goal for this person. This individual is signaling that they value highly the beliefs that they currently have and that they put a very small premium on mechanisms that might improve belief accuracy (but that involve belief change).

In short, thinking dispositions of the type we have examined provide information about epistemic goals at the rational level of analysis. Within such a conceptualization, we can perhaps better understand why the thinking dispositions predicted additional variance in AET performance after cognitive ability was partialled out. This result may indicate that to understand variation in reasoning in such a task we need to examine more than just differences at the algorithmic level (computational capacity)—we must know something about the epistemic goals of the reasoners.

In short, performance on the AET—a task requiring reasoning about previously held beliefs—is certainly dependent on the computational capacity of the participant, but it also depends on the balance of epistemic goals held by the reasoners. The instructions for the task dictate that one totally discount prior belief in evaluating the argument, but individuals may differ in their willingness or ability to adapt to such instructions. Individuals who put a low epistemic priority on maintaining current beliefs may find the instructions easy to follow. In contrast, those who attach high epistemic value to belief maintenance might find the standard instructions much harder to follow. Thus, epistemic goals might directly affect performance on such a task independent of the computational capacity that can be brought to bear to evaluate the argument. Some individuals may put a low priority on allocating computational capacity to evaluate the argument. Instead, for them, capacity is engaged to assess whether the conclusion is compatible with prior beliefs. Other individuals—of equal cognitive ability—may marshal their cognitive resources to decouple (see Navon, 1989a, 1989b) argument evaluation from their prior beliefs as the instructions demand. These individuals may easily engage in such a processing strategy because it does not conflict with their epistemic goals. Many problems in practical reasoning may have a similar structure (Baron, 1991, 1995; Foley, 1991; Davidson, 1995; Galotti, 1989; Harman, 1995; Jacobs & Potenza, 1991; Kardash & Scholes, 1996; Klaczynski & Gordon, 1996; Kuhn, 1991; Perkins, Farady, & Bushey, 1991). Such problems—although they obviously stress algorithmic capacity to varying degrees—might also have enormous variation in how they engage people's goal structure. To fully understand variation in human performance on such tasks, we might need to consider variation at the rational level as well as at the algorithmic level of cognitive analysis.

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14, 471–517.

- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (Ed.), *Advances in reading/language research* (Vol. 2, pp. 231–256). Greenwich, CT: JAI Press.
- Audi, R. (1993a). *Action, intention, and reason*. Ithaca, NY: Cornell University Press.
- Audi, R. (1993b). *The structure of justification*. Cambridge, England: Cambridge University Press.
- Babad, E. (1987). Wishful thinking and objectivity among sport fans. *Social Behavior: An International Journal of Applied Social Psychology*, 4, 231–240.
- Babad, E., & Katz, Y. (1991). Wishful thinking—Against all odds. *Journal of Applied Social Psychology*, 21, 1921–1938.
- Baron, J. (1985). *Rationality and intelligence*. Cambridge, England: Cambridge University Press.
- Baron, J. (1988). *Thinking and deciding*. Cambridge, England: Cambridge University Press.
- Baron, J. (1991). Beliefs about thinking. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 169–186). Hillsdale, NJ: Erlbaum.
- Baron, J. (1993). Why teach thinking?—An essay. *Applied Psychology: An International Review*, 42, 191–214.
- Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences*, 17, 1–42.
- Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning*, 1, 221–235.
- Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54, 569–579.
- Bell, D., Raiffa, H., & Tversky, A. (Eds.). (1988). *Decision making: Descriptive, normative, and prescriptive interactions*. Cambridge, England: Cambridge University Press.
- Bratman, M. E., Israel, D. J., & Pollack, M. E. (1991). Plans and resource-bounded practical reasoning. In J. Cummins & J. Pollock (Eds.), *Philosophy and AI: Essays at the interface* (pp. 7–22). Cambridge, MA: MIT Press.
- Brenner, L. A., Koehler, D. J., & Tversky, A. (1996). On the evaluation of one-sided evidence. *Journal of Behavioral Decision Making*, 9, 59–70.
- Bronfenbrenner, U., & Ceci, S. J. (1994). Nature–nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, 101, 568–586.
- Broniarczyk, S., & Alba, J. W. (1994). Theory versus data in prediction and correlation tasks. *Organizational Behavior and Human Decision Processes*, 57, 117–139.
- Brookfield, S. (1987). *Developing critical thinkers*. San Francisco, CA: Jossey-Bass.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Perspectives on socially shared cognition. *Educational Researcher*, 18(1), 32–42.
- Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119, 197–253.
- Campbell, D. T. (1987). Evolutionary epistemology. In G. Radnitzky & W. W. Bartley (Eds.), *Evolutionary epistemology, rationality, and the sociology of knowledge* (pp. 47–89). La Salle, IL: Open Court.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.
- Caryl, P. G. (1994). Early event-related potentials correlate with inspection time and intelligence. *Intelligence*, 18, 15–46.
- Ceci, S. J. (1990). *On intelligence . . . more or less: A bioecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice-Hall.
- Ceci, S. J. (1993). Contextual trends in intellectual development. *Developmental Review*, 13, 403–435.
- Cederblom, J. (1989). Willingness to reason and the identification of the self. In E. Maimon, D. Nodine, & O'Conner (Eds.), *Thinking, reasoning, and writing* (pp. 147–159). New York: Longman.
- Cherniak, C. (1986). *Minimal rationality*. Cambridge, MA: MIT Press.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooksey, R. W., & Freebody, P. (1987). Aspects of a computer-managed test of children's reading vocabulary: Reliability, validity, and characterization of knowledge. *Reading Psychology*, 8, 103–118.
- Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition*, 50, 41–77.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1–73.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory*. Odessa, FL: Psychological Assessment Resources.
- Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill-Palmer Quarterly*, 41, 328–346.
- Deary, I. J. (1995). Auditory inspection time and intelligence: What is the direction of causation? *Developmental Psychology*, 31, 237–250.
- Deary, I. J., & Stough, C. (1996). Intelligence and inspection time. *American Psychologist*, 51, 599–608.
- Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
- Detterman, D. K. (1994). Intelligence and the brain. In P. A. Vernon (Eds.), *The neuropsychology of individual differences* (pp. 35–57). San Diego, CA: Academic Press.
- Donaldson, M. (1978). *Children's minds*. London: Fontana Paperbacks.
- Donaldson, M. (1993). *Human minds: An exploration*. New York: Viking-Penguin.
- Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). New York: Freeman.
- Ennis, R. H., & Millman, J. (1985). *Cornell Critical Thinking Test*. Pacific Grove, CA: Midwest Publications.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49, 709–724.
- Epstein, S., & Meier, P. (1989). Constructive thinking: A broad coping variable with specific components. *Journal of Personality and Social Psychology*, 57, 332–350.
- Erwin, T. D. (1981). *Manual for the Scale of Intellectual Development*. Harrisonburg, VA: Developmental Analytics.
- Erwin, T. D. (1983). The Scale of Intellectual Development: Measuring Perry's scheme. *Journal of College Student Personnel*, 24, 6–12.

- Estes, W. K. (1982). Learning, memory, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 170–224). Cambridge, England: Cambridge University Press.
- Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*, 451–468.
- Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. London: Erlbaum Associates.
- Evans, J. St. B. T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, *11*, 295–306.
- Evans, J. St. B. T., Over, D. E., & Manktelow, K. (1993). Reasoning, decision making and rationality. *Cognition*, *49*, 165–187.
- Facione, P. (1992). *California Critical Thinking Dispositions Inventory*. La Cruz, CA: California Academic Press.
- Foley, R. (1991). Rationality, belief, and commitment. *Synthese*, *89*, 365–392.
- Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, *100*, 298–319.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*, 385–400.
- Galotti, K. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, *105*, 331–351.
- Gauthier, D. (1986). *Morals by agreement*. Oxford, England: Oxford University Press.
- Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Cambridge, MA: Harvard University Press.
- Goff, M., & Ackerman, P. L. (1992). Personality–intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, *84*, 537–552.
- Goldman, A. I. (1978). Epistemics: The regulative theory of cognition. *Journal of Philosophy*, *55*, 509–523.
- Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.
- Granberg, D., & Brent, E. (1983). When prophecy bends: The preference–expectation link in the U.S. presidential elections, 1952–1980. *Journal of Personality and Social Psychology*, *45*, 477–491.
- Harman, G. (1986). *Change in view*. Cambridge, MA: MIT Press.
- Harman, G. (1995). Rationality. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol 3, pp. 175–211). Cambridge, MA: MIT Press.
- Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review*, *85*, 109–130.
- Hunt, E. (1987). The next word on verbal ability. In P. A. Vernon (Ed.), *Speed of information-processing and intelligence* (pp. 347–392). Norwood, NJ: Ablex.
- Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development*, *62*, 166–178.
- Kardash, C. M., & Scholes, R. J. (1996). Effects of pre-existing beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology*, *88*, 260–271.
- Keating, D. P. (1990). Charting pathways to the development of expertise. *Educational Psychologist*, *25*, 243–267.
- Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston.
- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment*. San Francisco: Jossey-Bass.
- Kitchener, K., & Fischer, K. W. (1990). A skill approach to the development of reflective thinking. In D. Kuhn (Eds.), *Developmental perspectives on teaching and learning thinking skills* (pp. 48–62). Basel, Switzerland: Karger.
- Kitcher, P. (1993). *The advancement of science*. New York: Oxford University Press.
- Klaczynski, P. A. (1997). Bias in adolescents' everyday reasoning and its relationship with intellectual ability, personal theories, and self-serving motivation. *Developmental Psychology*, *33*, 273–283.
- Klaczynski, P. A., & Gordon, D. H. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology*, *62*, 317–339.
- Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, *56*, 28–55.
- Kornblith, H. (1993). *Inductive inference and its natural ground*. Cambridge, MA: MIT Press.
- Kramer, D. A., Kahlbaugh, P., & Goldston, R. (1992). A measure of paradigm beliefs about the social world. *Journal of Gerontology: Psychological Sciences*, *47*, P180–P189.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York: Plenum Press.
- Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing the mind: "Seizing" and "freezing." *Psychological Review*, *103*, 263–283.
- Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.
- Kuhn, D. (1993). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly*, *38*, 74–103.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Lipman, M. (1991). *Thinking in education*. Cambridge, England: Cambridge University Press.
- Lohman, D. F. (1989). Human intelligence: An introduction to advances in theory and research. *Review of Educational Research*, *59*, 333–373.
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*, 1231–1243.
- Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Cambridge, MA: Harvard University Press.
- Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory and Cognition*, *17*, 11–17.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.
- Moshman, D. (1994). Reasoning, metareasoning, and the promotion of rationality. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind, and reasoning: Structure and development* (pp. 135–150). Amsterdam: Elsevier.
- Navon, D. (1989a). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system: I. Basic tenets. *European Journal of Cognitive Psychology*, *1*, 191–213.
- Navon, D. (1989b). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system: II. Application to the field of attention. *European Journal of Cognitive Psychology*, *1*, 215–238.

- Neimark, E. (1987). *Adventures in thinking*. San Diego, CA: Harcourt Brace Jovanovich.
- Newell, A. (1982). The knowledge level. *Artificial Intelligence*, 18, 87-127.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Nickerson, R. (1987). Why teach thinking? In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 27-40). New York: Freeman.
- Norman, D. A. (1976). *Memory and attention: An introduction to human information processing*, (2nd ed.). New York: Wiley.
- Norris, S. P. (1992). Testing for the disposition to think critically. *Informal Logic*, 14, 157-164.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
- Nozick, R. (1993). *The nature of rationality*. Princeton, NJ: Princeton University Press.
- Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, 31, 117-140.
- Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 31-60). London: Routledge.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631.
- Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning*, 1, 121-152.
- Olson, D. R. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review*, 47, 257-281.
- Olson, D. R. (1986). The cognitive consequences of literacy. *Canadian Psychology*, 27, 109-121.
- Olson, D. R. (1994). *The world on paper*. Cambridge, England: Cambridge University Press.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317.
- Perkins, D. N. (1995). *Outsmarting IQ: The emerging science of learnable intelligence*. New York: Free Press.
- Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 83-105). Hillsdale, NJ: Erlbaum.
- Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly*, 39, 1-21.
- Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart & Winston.
- Piaget, J. (1926). *The language and thought of the child*. London: Routledge & Kegan Paul.
- Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development*, 15, 1-12.
- Pintrich, P. R., Marx, R., & Boyle, R. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167-199.
- Pylyshyn, Z. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 297-340). San Diego, CA: Academic Press.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes* (Vol. 1, pp. 560-564). San Diego, CA: Academic Press.
- Rogoff, B., & Lave, J. (Eds.). (1984). *Everyday cognition*. Cambridge, MA: Harvard University Press.
- Rokeach, M. (1960). *The open and closed mind*. New York: Basic Books.
- Schoemaker, P. (1991). The quest for optimality: A positive heuristic of science? *Behavioral and Brain Sciences*, 14, 205-245.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82, 498-504.
- Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology*, 85, 406-411.
- Schommer, M. (1994). Synthesizing epistemological belief research: Tentative understandings and provocative confusions. *Educational Psychology Review*, 6, 293-319.
- Schommer, M., & Walker, K. (1995). Are epistemological beliefs similar across domains? *Journal of Educational Psychology*, 87, 424-432.
- Siegel, H. (1993). Not by skill alone: The centrality of character to critical thinking. *Informal Logic*, 15, 163-177.
- Slusher, M. P., & Anderson, C. A. (1996). Using causal persuasive arguments to change beliefs and teach new information: The mediating role of explanation availability and evaluation bias in the acceptance of knowledge. *Journal of Educational Psychology*, 88, 110-122.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.
- Stanovich, K. E. (1989). Implicit philosophies of mind: The Dualism Scale and its relationships with religiosity and belief in extrasensory perception. *Journal of Psychology*, 123, 5-23.
- Stanovich, K. E. (1994). Reconceptualizing intelligence: Dysrationalia as an intuition pump. *Educational Researcher*, 23(4), 11-22.
- Stanovich, K. E. (in press). *Variable rationality*. Mahwah, NJ: Erlbaum.
- Sternberg, R. J., & Ruzgis, P. (Eds.). (1994). *Personality and intelligence*. Cambridge, England: Cambridge University Press.
- Straughn, C. S., & Straughn, B. L. (Eds.). (1995). *Lovejoy's college guide* (23rd ed.). New York: Simon & Schuster.
- Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.
- Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena. *Journal of Personality and Social Psychology*, 44, 1029-1037.
- Troidahl, V., & Powell, F. (1965). A short-form dogmatism scale for use in field studies. *Social Forces*, 44, 211-215.
- Vernon, P. A. (1991). The use of biological measures to estimate behavioral intelligence. *Educational Psychologist*, 25, 293-304.
- Vernon, P. A. (1993). *Biological approaches to the study of human intelligence*. Norwood, NJ: Ablex.
- Vernon, P. A., & Mori, M. (1992). Intelligence, reaction times, and

- peripheral nerve conduction velocity. *Intelligence*, 16, 273–288.
- Voss, J., Perkins, D. & Segal, J. (Eds.). (1991). *Informal reasoning and education*. Hillsdale, NJ: Erlbaum.
- Wade, C., & Tavis, C. (1993). *Critical and creative thinking*. New York: Harper Collins.
- Watson, G., & Glaser, E. M. (1980). *Watson-Glaser Critical Thinking Appraisal*. New York: Psychological Corporation.
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, 67, 1049–1062.
- West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science*, 2, 325–330.
- White, T. G., Slater, W. H., & Graves, M. F. (1989). Yes/no method of vocabulary assessment: Valid for whom and useful for what? In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction* (38th yearbook of the National Reading Conference, pp. 391–397). Chicago, IL: National Reading Conference.
- Zechmeister, E. B., & Johnson, J. (1992). *Critical thinking: A functional approach*. Pacific Grove, CA: Brooks/Cole.
- Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. (1977). A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition. *Intelligence*, 1, 5–31.

Appendix A

Examples of Items on the Argument Evaluation Test

Instructions: We are interested in your ability to evaluate counter-arguments. First, you will be presented with a belief held by an individual named Dale. Following this, you will be presented with Dale's premise or justification for holding this particular belief. A Critic will then offer a counter-argument to Dale's justification for the belief. (We will assume that the Critic's statement is factually correct.) Finally, Dale will offer a rebuttal to the Critic's counter-argument. (We will assume that Dale's rebuttal is also factually correct.) You are to evaluate the strength of Dale's rebuttal to the Critic's counter-argument, regardless of your feeling about the original belief or Dale's premise.

Dale's belief: The national debt should be reduced by cutting Congressional salaries.

Dale's premise or justification for belief: Congressional salaries are very high, and cutting them would make a significant step towards paying off the huge national debt.

Critic's counter-argument: The national debt is so large that totally eliminating all Congressional salaries would still hardly make a dent in it (assume statement factually correct).

Dale's rebuttal to Critic's counter-argument: The members of Congress, whose actions are to a considerable extent responsible for the huge national debt, earn salaries several times higher than the national average (assume statement factually correct).

Indicate the strength of Dale's rebuttal to the Critic's counter-argument:

1 = Very Weak 2 = Weak 3 = Strong 4 = Very Strong

Median experts' evaluation (Argument Quality): 1.0

Mean item agreement across participants (Prior Belief): 3.00

Mean rebuttal evaluation across participants (Argument Evaluation): 2.70

Dale's belief: The present Social Security system is unfair to people who are now retired.

Dale's premise or justification for belief: The present Social Security system is unfair to retired people because these people get far too little money.

Critic's counter-argument: Currently, retirees are drawing from the Social Security system over four times what they con-

tributed to the system before they die (assume statement factually correct).

Dale's rebuttal to Critic's counter-argument: These figures don't take into account the fact that retirees are being paid in dollars that have been greatly eroded over the years due to inflation (assume statement factually correct).

Indicate the strength of Dale's rebuttal to the Critic's counter-argument:

1 = Very Weak 2 = Weak 3 = Strong 4 = Very Strong

Median experts' evaluation (Argument Quality): 3.0

Mean item agreement across participants (Prior Belief): 2.36

Mean rebuttal evaluation across participants (Argument Evaluation): 2.89

Dale's belief: Students should have a stronger voice than the general public in setting university policies.

Dale's premise or justification for belief: Because students are the ones who must ultimately pay the costs of running the university through tuition, they should have a stronger voice in setting university policies.

Critic's counter-argument: Tuition covers less than one half the cost of an education at most state universities (assume statement factually correct), so the taxpayers should have a stronger say in the policies.

Dale's rebuttal to Critic's counter-argument: Because it is the students who are directly influenced by university policies (assume statement factually correct), they are the ones who should have the stronger voice.

Indicate the strength of Dale's rebuttal to the Critic's counter-argument:

1 = Very Weak 2 = Weak 3 = Strong 4 = Very Strong

Median experts' evaluation (Argument Quality): 1.0

Mean item agreement across participants (Prior Belief): 3.32

Mean rebuttal evaluation across participants (Argument Evaluation): 3.11

Appendix B

Intercorrelations Among the Primary Variables

Variable	1	2	3	4	5	6	7
1. Flexible thinking	—						
2. Openness-ideas	.36	—					
3. Openness-values	.36	.40	—				
4. Absolutism	-.37	-.39	-.59	—			
5. Dogmatism	-.19	-.22	-.54	.55	—		
6. Categorical thinking	-.31	-.31	-.58	.57	.54	—	
7. Superstitious thinking	-.26	-.13	-.13	.23	.19	.28	—
8. Counterfactual thinking	.06	-.02	.09	-.09	-.15	-.08	.01

Note. Correlations larger than .11 are significant at the .05 level (two-tailed).

Appendix C

Component Loadings for All Variables After Varimax Rotation

Variable	Component	
	1	2
Flexible thinking	.566	-.369
Openness-ideas	.568	-.355
Openness-values	.805	—
Absolutism	.816	—
Dogmatism	.721	—
Categorical thinking	.790	—
Superstitious thinking	-.379	.381
Counterfactual thinking	—	.735
% variance accounted for	40.8	13.5

Note. Component loadings lower than .350 have been eliminated.

Received March 15, 1996

Revision received July 16, 1996

Accepted August 28, 1996 ■