Journal of Educational Psychology
1999, Vol. 91, No. 3, 497–510

0022-0663/99/$3.00

The Domain Specificity and Generality of Belief Bias: Searching for a Generalizable Critical Thinking Skill

Walter C. Sá
University of Toronto

Richard F. West
James Madison University

Keith E. Stanovich
University of Toronto

The domain specificity and generality of belief-biased reasoning was examined across a height judgment task and a syllogistic reasoning task that differed greatly in cognitive requirements. Moderate correlations between belief-bias indices on these 2 tasks falsified an extreme form of the domain specificity view of critical thinking skills. Two measures of cognitive ability and 2 measures of cognitive decontextualization skill were positively correlated with belief bias in a height judgment task where prior knowledge accurately reflected an aspect of the environment and negatively correlated with belief bias in a height judgment task where prior knowledge was incongruent with the environment. Likewise, cognitive ability was associated with skill at resisting the influence of prior knowledge in the syllogistic reasoning task. Participants high in cognitive ability were able to flexibly use prior knowledge, depending upon its efficacy in a particular environment. They were more likely to project a relationship when it reflected a useful cue, but they were also less likely to project a prior belief when the belief was inefficacious.

An important research tradition within cognitive science has demonstrated how prior belief biases the evaluation of arguments and of data (Baron, 1995; Evans, Over, & Manktelow, 1993; George, 1995; Klaczynski, Gordon, & Fauth, 1997; Klaczynski & Narasimham, 1998; Moshman & Franks, 1986). The quintessential paradigm for demonstrating this phenomenon is the syllogistic reasoning problem, which pits the believability of the conclusion against the validity of the argument (e.g., Evans, Barston, & Pollard, 1983; Markovits & Nantel, 1989). In this paradigm, the belief bias effect occurs when participants are found to judge the validity of the syllogism more accurately when the believability of the conclusion coincides with the validity of the syllogism than when it conflicts. For example, problems that are invalid and have unbelievable conclusions (e.g., All guns are dangerous. Rattlesnakes are dangerous. Therefore, rattlesnakes are guns.) are easier than problems that are logically invalid and have believable conclusions (e.g., All living things need water. Roses need water. Roses are living things.). Presumably, in the latter situation, the evaluation of logical validity is disrupted by the real-world knowledge that is cuing an alternative response. Prior belief (in this case, knowledge of the world) thus interferes with optimal task performance, which is attained by attending only to logical validity.

Belief bias based on prior knowledge of real-world relationships has also been demonstrated in other paradigms such as the evaluation of numerical covariation information (Broniarczyk & Alba, 1994). Additionally, belief bias, based not on prior knowledge but on opinion about a controversial issue, has been demonstrated in several paradigms such as covariation detection (Levin, Wasserman, & Kao, 1993; Stanovich & West, 1998b), argument evaluation (Klaczynski, 1997; Stanovich & West, 1997, 1998a), and evidence evaluation (Kardash & Scholes, 1996; Klaczynski & Gordon, 1996; Klaczynski et al., 1997; Klaczynski & Narasimham, 1998; Kuhn, 1991; Slusher & Anderson, 1996). Similarly, the social psychological literature contains many demonstrations of how the evaluation of communications is biased by prior opinion and belief (e.g., Biek, Wood, & Chaiken, 1996; Dole & Sinatra, 1998; Kunda, 1990; Lord, Lepper, & Preston, 1984; Nickerson, 1998; Wegener & Petty, 1997). Belief biases of these various types exemplify what Stanovich (1999) has termed the fundamental computation bias of human cognition, which refers to a bias that arises because of difficulties in cognitive decontextualization.

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

Walter C. Sá is now at the Department of Psychology, Grand Valley State University.

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada. Preparation of the manuscript was supported by a Connaught Research Fellowship in the Social Sciences, awarded to Keith E. Stanovich.

We thank Monica Biernat for generously providing the photographs from her 1990 study. We also thank Michael Doherty and Denis Hilton for their comments on an earlier version of the manuscript.

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. Electronic mail may be sent to kstanovich @oise.utoronto.ca.

As in the syllogistic reasoning literature, all of these effects represent cases where prior knowledge, belief, and opinion serve to disrupt the impartial evaluation of evidence and argument. The ability to avoid this type of belief bias is repeatedly stressed as a positive trait in the literature on critical thinking. For example, Norris and Ennis (1989) argued that one important characteristic of critical thinking is the tendency to "reason from starting points with which they disagree without letting the disagreement interfere with 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) 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" (p. 30). This sentiment has been echoed by many critical thinking theorists (e.g., Baron, 1991, 1995; Baron & Sternberg, 1987; Klaczynski & Narasimham, 1998; Kuhn, 1991, 1996; Moshman, 1994; Perkins, 1995; Perkins, Jay, & Tishman, 1993; Siegel, 1988, 1997).

In these statements championing the avoidance of belief-biased reasoning, the ability to evaluate evidence in an unbiased manner has been treated as a global trait. For example, when Nickerson (1987) warned against "differentially weighting evidence according to personal preferences" (p. 30), the domain of the preferences is not specified. Likewise, when Zechmeister and Johnson (1992) championed the ability to "accept statements as true even when they don't agree with one's own position" (p. 6), they did not specify a statement domain. Throughout the entire critical thinking literature, domain generality has been assumed for the various thinking styles that are listed as the defining features of critical thinking (e.g., Ennis, 1987; Lipman, 1991; Wade & Tavris, 1993). Indeed, Baron (1985b) made some degree of domain generality a defining feature of his notion of a thinking style: "Cognitive styles ought to be general. By ought I mean that evidence against the generality of a style is taken to make the style less interesting" (pp. 379–380). This view leads to the obvious individual difference prediction that "we should expect some correlation across individuals between style in one situation and style in another, regardless of how discrepant the situations are" (Baron, 1985b, p. 380). However, there have been virtually no such multivariate studies reported on the belief bias effect. Whether the phenomenon has any degree of domain generality or whether it is domain specific (and thus, under Baron's criterion, does not warrant its treatment as a thinking style in the critical thinking literature) is almost completely unknown.

In contrast to the assumption of domain generality in the critical thinking literature, in developmental and cognitive psychology, the reigning assumption for at least a decade has been one of domain specificity. For example, the contextualist tradition within developmental psychology emphasizes the point that the exercise of cognitive skills is often quite situation-specific (J. S. Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991; Rogoff & Lave, 1984). Likewise, theorists who have emphasized the importance of domain knowledge (e.g., Alexander, Kulikowich, & Schulze, 1994; Ceci, 1993, 1996) have argued that many so-called basic cognitive processes are so dependent on familiarity with the specific stimulus domain and its context that it seems almost a misnomer to call them "basic" (see Ceci, 1996). Stigler and Baranes (1988) stressed this point in particular by arguing that "it is more and more difficult, within modern cognitive theory, to draw a division between basic processes and 'mere' content" (p. 257).

Thus, theorists in both the domain-knowledge tradition and the contextualist tradition of developmental and educational psychology emphasize the domain specificity of the exercise of a cognitive skill or style (for a critique of this view, see J. R. Anderson, Reder, & Simon, 1996). Such theorists question the existence of thinking styles that have the generality ascribed by Baron (1985b). For example, under a contextualist conceptualization, belief bias effects would be expected to display extreme domain specificity. General tendencies to avoid belief bias, such as those discussed in the critical thinking literature, are thus treated with extreme skepticism within the contextualist framework.

That there are influential but competing traditions within which belief bias effects can be conceptualized (the critical thinking literature, the contextualist tradition, and the domain-knowledge tradition) highlights the urgency of producing empirical data that can at least partially adjudicate the differential predictions. In the present article, we present one of the few attempts to examine the domain generality of belief bias as a cognitive style cutting across tasks that widely differ in cognitive and response requirements.

Research Strategy

In the present study, we compared two tasks on which belief bias has been displayed. One task, typical of verbal reasoning paradigms, is the syllogistic reasoning task discussed previously. The second task was drawn from the perceptual judgment domain in order to contrast with the verbal reasoning domain of the syllogistic reasoning task. This task was adapted from the work of Nelson, Biernat, and Manis (1990) who had participants judge the heights of seated and standing males and females in photographs. They found that the judgments were related to the actual heights, but that the over- and underestimations of the actual heights were related to the gender of the person in the target photograph. That is, although gender is a valid cue in that it is related to actual height, participants tended to overproject this cue. Importantly, this over projection was maintained in a condition (termed the *matched condition*) where the participants were (a) informed that the male and female pictures were matched in height, and (b) warned not to use gender as a cue.

From the standpoint of the present investigation, the matched condition of the Nelson et al. (1990) height judgment task is important because it mirrors the logic of the syllogistic reasoning task. In both cases, the participant is told to ignore a cue that is normally diagnostic but in this particular situation is not predictive (the truth status of the conclusion in the syllogistic reasoning task and the gender of the stimulus in the height judgment task). In both cases, it has been found that, in the aggregate, people are unable to ignore the nondiagnostic cue and evaluate only the information of interest (the validity of the syllogism and the valid height cues in the photograph, respectively). These tasks are particularly interesting for investigating issues of domain specificity and generality because, despite the similarity of the belief bias logic across the two, the stimuli used in the tasks are very different, and the judgments are also quite different (one is a judgment of logical validity and the other a height judgment).

Thus, in the present investigation, in addition to a syllogistic reasoning task, we examined two versions of the Nelson et al. (1990) height judgment paradigm. We compared a situation where ecologically valid stimuli were used with one where the stimuli that were chosen conflicted with prior belief. In the former—termed the *ecological condition*—the photographs viewed by the participants were a reasonably representative sample of the actual heights of males and females in North America, and participants were informed of this fact. In the latter condition—the matched condition—participants viewed a sample of photographs where the mean heights of the males and females were matched, and participants were informed of the matching (that is, they were informed that the gender cue was nondiagnostic). This condition is the one that is structurally analogous to the syllogistic reasoning conflict condition.

In our study questionnaire, we included measures of styles of epistemic regulation that embody the fundamental assumption of the critical thinking literature—that actively open-minded thinking is exercised in a domain-general manner (Baron, 1993, 1994; Perkins et al., 1993). We constructed our questionnaire by drawing upon scales already in the literature and by constructing some of our own. Several of our measures had strong similarities to other related measures in the literature (see Klaczynski et al., 1997; Schommer, 1990, 1993, 1994; Schommer & Walker, 1995; Webster & Kruglanski, 1994). Perhaps the strongest similarities are with the two factors that Schommer (1990, 1993) called belief in simple knowledge and belief in certain knowledge. That is, we particularly focused on thinking styles that had potential epistemic significance (see Cederblom, 1989); 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, 1985a, p. 15). We attempted to tap the following dimensions: epistemological absolutism, willingness to switch perspective, willingness to decontextualize, identification with beliefs, and the tendency to consider alternative opinions and evidence. The focus with respect to these measures was on whether they predicted the actual magnitude of the belief bias effect and whether that predictive validity was domain general.

We included converging measures of cognitive ability in order to examine the type of processing to which the human cognitive apparatus is more prone when freed of computational limitations. If the received position in the critical thinking literature is correct, people with greater cognitive and reasoning ability should show a greater tendency to evaluate evidence and arguments independent of prior belief (i.e., they should display the cognitive flexibility necessary to detach reasoning from prior knowledge). In contrast, if contextualist theories are correct, tendencies toward decontextualization might be viewed as inefficacious, and people with increased cognitive ability might display greater reluctance to detach world knowledge from their information processing (i.e., they will demonstrate increased levels of belief bias). Finally, we included two other tasks that require cognitive decontextualization in order to test the domain generality of styles of epistemic regulation.

Method

Participants

The participants, who consisted of 124 students (54 males and 70 females), were recruited at a large Canadian university through poster advertisements that were distributed on campus. Their mean age was 21.8 years ($SD = 3.7$), and the modal age was 19. Participants were paid $20 per student for their cooperation.

Belief Bias Tasks

Height judgment task. The height judgment task was adapted from the paradigm used by Nelson et al. (1990). Participants were presented with full-length photographs of males and females, who were seated in a variety of natural settings. The participants' task was to judge as accurately as possible the heights of the individuals in each of these photographs. The stimuli in the first height judgment task—termed the *ecological height judgment task*—were sampled to reflect the actual male and female heights in the population. These stimuli consisted of 83 pictured models, 40 males and 43 females. Seventy-five of these came from the Nelson et al. (1990) study and were generously provided by those investigators. The mean height of the males in the photographs was 5 ft, 10.6 in. ($SD = 2.7$ in.), and the mean height of the females in the photographs was 5 ft, 5.6 in. ($SD = 3.4$ in.). These mean heights are roughly equal to those of the general population to which the participants belonged (National Dairy Council, 1980). The photographs were taken at varying distances. Thus, there was no significant correlation between image size and actual height ($r = -.159$).

The photographs were presented in a randomized order, mixing males and females. The instructions in the ecological height judgment task were as follows:

> This task is designed to assess how accurately people can judge the physical characteristics of individuals based on a small amount of information. You will be looking at photographs of individuals and will be asked to judge the height of each person pictured. The task is fairly difficult because the models are all seated. On each page of this booklet you will find a picture of a person. Your task is simply to look at each photograph and estimate the height of the pictured person. Each person's height was measured when they were wearing the shoes that they are wearing in the picture. Therefore, your height estimate should be an estimate of the person's height

> while wearing the shoes in the picture. As you flip through these pictures, read out loud to the experimenter your height estimate.

Participants were allowed to give their estimates either in centimeters or in feet and inches (the vast majority preferred the latter). The judgments of the few participants who gave centimeter estimates were transformed into feet and inches.

The second version of the height judgment task—termed the *matched height judgment task*—immediately followed the ecological version. In this version, participants were specifically informed that the gender cue had been rendered nondiagnostic and consequently should be ignored in their ensuing height judgments. Participants estimated the heights of 46 pictured models (23 male and 23 female), where the mean male and female heights were equated at 5 ft, 8.5 in. (*SD*s = 3.4 and 3.6, respectively). Again, the stimuli came from the Nelson et al. (1990) study and were presented in a randomized order, mixing males and females. All of these 46 stimuli were different from those used in the ecological condition. The instructions in the matched height judgment task were as follows:

> In the following task, you will be doing exactly what you were just doing before. You will again be looking at several pictured people and estimating their height. This time, however, the men and women pictured are, on average, of *equal height.* That is, we have taken care to match the heights of the men and women pictured (as before, footwear is included in the measurement). For every woman of a particular height, somewhere in this section there is also a man of that same height. Therefore, in order to make as accurate a height judgment as possible, try to judge each photograph as an individual case. *Do not rely on the person's sex to aid your judgment.*

One participant failed to complete these tasks and 6 were removed as multivariate outliers (having Mahalanobis distance metrics greater than 4.0; see Tabachnick & Fidell, 1983).

Syllogistic reasoning task. Twenty-four syllogistic reasoning problems, largely drawn from Markovits and Nantel (1989), were completed by the participants. Eight of the problems were worded such that the validity judgment was in conflict with the believability of the conclusion (e.g., All living things need water; Roses need water; therefore, Roses are living things—which is invalid). These were termed the *inconsistent items.* Eight of the problems were worded such that the validity judgment was congruent with the believability of the conclusion (e.g., All fish can swim; Tuna are fish; therefore, Tuna can swim—which is valid). These were termed the *consistent items.* Eight of the problems involved imaginary content (e.g., All opprobines run on electricity; Jamtops run on electricity; therefore, Jamtops are opprobines—which is invalid). These were termed the *neutral items.* The instructions that were given to the participants were as follows:

> In the following problems, you will be given two premises *which you must assume are true.* A conclusion from the premises then follows. You must decide whether the conclusion *follows logically* from the premises or not. You must *suppose that the premises are all true* and limit yourself only to the information contained in the premises. This is very important. Decide if the conclusion follows logically from the premises, assuming the premises are true, and circle your response.

After each item, the participants indicated their responses by circling one of two alternatives: (a) "Follows Logically," or (b) "Does Not Follow Logically."

Cognitive Decontextualization Tasks

Covariation judgment. For this task, we adapted a paradigm where people were presented with covariation information that was accommodated by the format of a 2 × 2 contingency table (see Schustack & Sternberg, 1981; Wasserman, Dorner, & Kao, 1990). Participants indicated their opinions on 25 sets of hypothetical relationships covering a variety of different topics (e.g., that couples who live together before marriage tend to have successful marriages) and then evaluated the degree of association between the same 25 sets of two variables in the data of a hypothetical research study. These data corresponded to the four cells of the 2 × 2 contingency table, traditionally labeled A, B, C, and D (see Levin et al., 1993). For example, in the "living together" problem, participants were told that a researcher had sampled 225 couples and had found that (a) 100 couples lived together and had successful marriages, (b) 50 couples lived together and were divorced, (c) 25 couples did not live together and had successful marriages, and (d) 50 couples did not live together and were divorced. Subsequent to the presentation of one set of these data, the participants were asked to judge the nature and extent of the relationship between living together before marriage and successful marriages in these data on a scale ranging from +10 (*positive association*) to −10 (*negative association*) and centered on 0 (*no association*). Details of the values used in the tables were presented in Stanovich and West (1998b).

The normatively appropriate strategy in this task (see Allan, 1980; Kao & Wasserman, 1993; Shanks, 1995) is to use the conditional probability rule (i.e., subtracting the probability of the target hypothesis when the indicator is absent from the probability of the target hypothesis when the indicator is present. Numerically, the rule amounts to calculating the Δp statistic: $[A/(A + B)] - [C/(C + D)]$, (see Allan, 1980). Thus, each participant's 25 covariation scores were regressed simultaneously on the 25 Δp values and the 25 prior-opinion scores. The former beta weight was used as the primary indicator of the ability to reason about covariation independently of their prior opinion on the issue in question.

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

Cognitive Ability Measures

Cognitive ability 1. Participants completed the Vocabulary and Block Design subtests of the Wechsler Adult Intelligence Scale–Revised (WAIS-R; Wechsler, 1981). Because these two subtests attain moderate (Block Design) and high (Vocabulary) correlations with the full-scale IQ score, their scores can be prorated to obtain a popular short-form IQ score with high reliability and validity (Sattler, 1988). Using Sattler's (1988) formulas, we derived a prorated IQ score for each participant, which we termed *general cognitive ability 1* (CA1). The mean of this variable in the sample was 108.3 (*SD* = 11.7).

Cognitive ability 2. The second cognitive ability measure, *cognitive ability 2* (CA2), was also derived by combining performance on a verbal and nonverbal test. The latter consisted of 18 problems from Raven's Advanced Progressive Matrices (Set II; Raven, 1962). The participants were given 15 min to complete the 18 items on the test. By eliminating 12 of the easiest problems, where performance in a university sample is near ceiling (Carpenter, Just, & Shell, 1990; Raven, Court, & Raven, 1977) and six of the most difficult problems where performance is nearly floored (Carpenter et al., 1990; Raven et al., 1977), a cut-time version of the advanced matrices was created with nearly the same reliability as the full test (Cahan & Cohen, 1989; Stanovich & Cunningham, 1992).

The verbal measure was a brief vocabulary measure (again, 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 a reliable and valid way of assessing individual differences in vocabulary knowledge (R. C. Anderson & Freebody, 1983; Stanovich, West, & Harrison, 1995). The stimuli for the task were 40 words and 20 pronounceable nonwords taken largely from the stimulus list of Zimmerman, Broder, Shaughnessy, & Underwood (1977). The words and nonwords were intermixed according to alphabetical order. 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.

The CA2 index was created by standardizing the raw scores on the Raven matrices and the vocabulary checklist measure and then adding the two standard scores together. The correlation between the two composite indices, CA1 and CA2, was .707.

Actively Open-Minded Thinking Questionnaire

Participants completed a questionnaire consisting of a number of subscales assessing styles of epistemic regulation. The following response format was used for each item in the questionnaire: 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, both with each other and with other scales that were not part of the present investigation. Most of the subscales were described in more detail in Stanovich and West (1997). Brief descriptions follow.

Flexible thinking scale. We devised the items on the flexible thinking scale, which were validated in a previous investigation (Stanovich & West, 1997). Items tap flexible thinking as a multifaceted construct that encompasses the cultivation of reflectiveness rather than impulsivity (e.g., "If I think longer about a problem I will be more likely to solve it"), the seeking and processing of information that disconfirms one's belief (e.g., "People should always take into consideration evidence that goes against their beliefs"), and the willingness to change one's beliefs in the face of contradictory evidence (see Baron, 1993, 1994). There were 10 items on the scale. Because this subscale is relatively new, the items are reproduced in Appendix A.

Openness–ideas. 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," and "I find philosophical arguments boring"—the latter was reverse scored).

Openness–values. 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," and "I believe letting students hear controversial speakers can only confuse and mislead them"—the latter was reverse scored).

Absolutism. This scale consisted of nine items adapted from the Scale of Intellectual Development (SID), which was developed by Erwin (1981, 1983). The SID represents an attempt to develop a multiple-choice scale to measure the early stages of Perry's (1970) model of intellectual development in young adulthood, which are characterized by cognitive rigidity, by a belief that issues can be couched in either–or terms, that there is one right answer to every complex problem, and by reliance on authority for belief justification (e.g., "It is better to simply believe in a religion than to be confused by doubts about it").

Dogmatism. The dogmatism subscale consisted of nine items. Three were taken from a short-form field version (Troldahl & Powell, 1965) of Rokeach's (1960) dogmatism scale, two from Paulhus and Reid (1991), and four from the full Rokeach scale published in Robinson, Shaver, and Wrightsman (1991). A typical item is, "No one can talk me out of something I know is right."

Categorical thinking. The following 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."

Belief identification. The belief identification scale was inspired by a theoretical paper by Cederblom (1989) in which he argued for a potential thinking style centered around the extent to which people identify their beliefs with their self-concept (e.g., "It makes me happy and proud when someone famous holds the same beliefs that I do"). A nine-item scale was distilled from Cederblom's discussion of this concept. Because this subscale is relatively new, the items are reproduced in Appendix A.

Need for closure. Twelve items taken from the Need for Closure Scale published in Kruglanski, Webster, and Klem (1993) were administered. A typical item is "I dislike it when a person's statement could mean many different things."

Procedure

Participants completed the tasks during a single 3–4 hr session in which they also completed some other tasks that were not part of the present investigation. All participants were individually tested by the same experimenter. The order of tasks completed was as follows: actively open-minded thinking questionnaire, syllogisms, height judgment task, WAIS–R subtests, AET, vocabulary checklist, covariation data evaluation, and Raven matrices.

Results

Projection of the Gender Cue in the Height Judgment Task

One index of the potency of the gender cue in the judgments of each participant is the correlation between the gender of the target in the photograph (G) and that participant's estimates, *r*(G, E). The mean value of this correlation in the ecological set was .693 (*SD* = .129), and the mean value of r(G, E) in the matched set was .281 (SD = .196)—a difference that was statistically significant, $t(116) = 26.37$, $p < .001$. The actual correlation between gender and the target heights in the photographs in the ecological set, r(G, T), was .628. Thus, the mean r(G, E) value of .693 in the ecological set represents only slightly more projection of the gender cue than is actually warranted. In the matched condition, the r(G, T) correlation was deliberately constrained to zero (actually, .018), therefore the mean r(G, E) value of .281, which is significantly greater than zero, $t(116) = 15.48$, $p < .001$, reliably represents more projection of the gender cue in this condition than is warranted.

Another metric that captures the degree of projection of the gender cue is simply the mean difference between the estimates of male targets and female targets for each participant. The mean difference in the ecological condition, $(M - F)_{\text{ecological}}$, ranged from 1.53 to 9.37 in. and averaged 4.13 in. (SD = 1.25). This mean is almost one inch less than the actual mean difference in the ecological set (5 in.). The mean difference in the matched condition, $(M - F)_{\text{matched}}$, was substantially lower (1.21, SD = 0.91), but it was still significantly different from zero, $t(116) = 14.35$, $p < .001$. In the matched condition, despite the instructions emphasizing that the males and females were matched in height, 111 of 117 participants gave mean male height estimates that were larger than their female estimates.

Accuracy in the Height Judgment Task: The Efficacy and Inefficacy of Projecting Prior Belief

There are two indices of accuracy in the height judgment task, each reflecting a different property of estimation efficacy (see N. R. Brown & Siegler, 1993). One index is the correlation between the estimates (E) and the target heights of the actual models (T). This correlation, r(T, E), was calculated for each participant and ranged from .395 to .766 in the ecological set and from .091 to .645 in the matched set. The mean value of r(T, E) was considerably higher in the ecological set (.610, SD = .071) than in the matched set (.358, SD = .109), $t(116) = 20.34$, $p < .001$.

The r(T, E) accuracy index reflects the tendency to *order* the targets correctly but it does not reflect the tendency to use the scale correctly—that is, to arrive at estimates that actually match the target heights. The sum of the absolute deviations (SAD) between estimate and target across all stimuli reflects this property. The smaller this index, the higher the absolute accuracy of the estimates. In principle, these two indices can be independent. For example, it is possible for r(T, E) to be quite high (indicating highly accurate ordering), and for SAD to also be high (indicating poor absolute accuracy of the estimates, i.e., poor scale use). This might occur if someone ordered the photographs nearly perfectly but had each of the estimates 3–4 inches too tall (because the individual believed that, overall, the population is taller than it is).

We might expect that in the ecological set, projecting one's knowledge of the gender relationship would be efficacious because in that set of stimuli there actually is a correlation between gender and target height, r(G, T) = .628. We found support for this conjecture. Using r(G, E) as a measure of the projection of this prior belief, we found that this index correlated highly with both r(T, E), $r = .697$, $p < .001$, and SAD, $r = -.525$, $p < .001$. That is, the more the participants projected the gender–height relationship, the more accurate they were. Participants who more strongly projected the gender–height relationship in their estimates tended to order the stimuli more accurately, and they made estimates that tracked the actual heights more closely in an absolute sense.

In the ecological set, even the tendency to overuse the gender cue (in short, to stereotype) is efficacious. This point is made clear in a subsequent analysis. Indicated in the top half of Figure 1 (analogous to the figures in Nelson et al., 1990), is the outcome when the aggregate height estimates (averaged over all 117 respondents) are regressed on gender and the actual height of the targets (additional regression analyses were run with image size as an additional predictor, but this variable had virtually no effect on this or any subsequent analysis, primarily because it had negligible correlations with gender, .005, actual height, −.159, and estimated height, −.018). The highly significant standard-

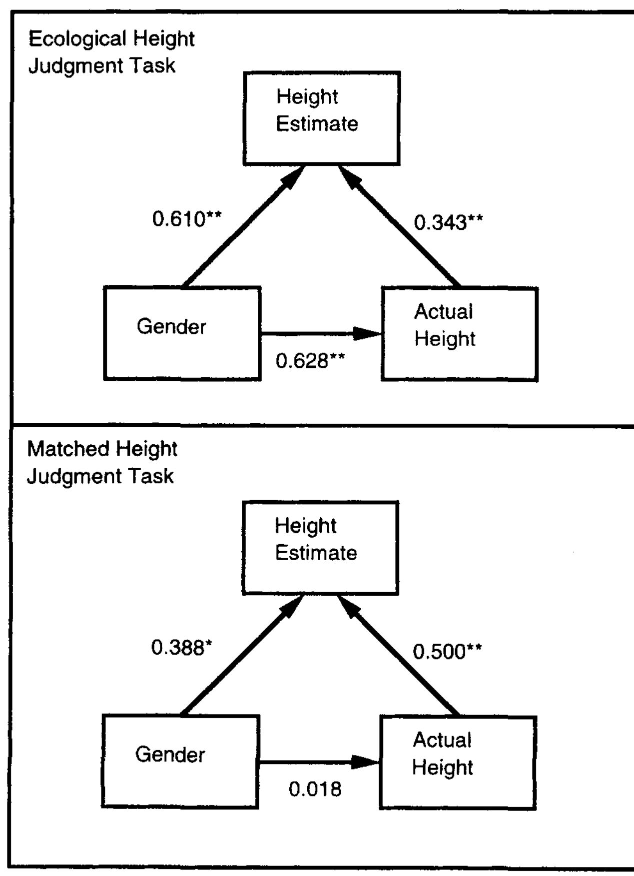

Figure 1. Top: Path diagram predicting the aggregate height judgments in the ecological set. Bottom: Path diagram predicting the aggregate height judgments in the matched set. Numbers on arrows are standardized beta weights.
*$p < .01$. **$p < .001$.

ized beta weight for gender is an indication that, in the words of Nelson et al. (1990), "male targets were judged to be taller, on average, than the female targets, even after the actual difference in height between the male and female targets had been statistically controlled" (p. 665). However, the tendency to stereotype on the basis of gender is actually associated with more accurate estimates. This analysis was carried out by running the regression analysis pictured in the top half of Figure 1 on each individual participant's scores and estimating 117 separate beta weights for gender and actual height (one pair for each participant). The former parameter was positively correlated with r(T, E), $r = .443$, $p < .001$, and negatively correlated with SAD, $r = -.314$, $p < .001$. Thus, the more individuals judged male targets to be taller, even after actual differences in male and female targets were statistically controlled, the more accurate they were, both in correctly ordering the stimuli and in their absolute deviations from the actual heights.

Of course, in the matched set, projection of the gender cue is not efficacious. There, the projection index, r(G, E), did not correlate significantly with either r(T, E), $r = -.172$, or SAD, $r = .143$—the more the participants projected the gender–height relationship, the less accurate they were. In the analysis that is summarized in the bottom half of Figure 1, aggregate height estimates (averaged over all 117 respondents) were regressed on gender and the actual height of the targets in the matched set. The standardized beta weight for gender—although reduced in magnitude from that obtained with the ecological set—was highly significant. Despite the instructions emphasizing that these targets were matched for height across gender, participants persisted in estimating greater male heights. Here, unlike the case in the ecological set, the tendency to stereotype on the basis of gender was not associated with more accurate estimates in the matched set. The regression analysis pictured in Figure 1 was run on each individual participant, and 117 separate beta weights for gender and actual height (one pair for each participant) were estimated. The former parameter failed to correlate positively with r(T, E), $r = -.181$, $p < .05$, or negatively with SAD, $r = .148$, *ns*. In fact, the signs of the correlations were in the opposite direction, significantly so in the former case—more stereotyped responders ordered the photographs less accurately.

Performance on the Syllogistic Reasoning Task

The mean number of correct responses on the syllogistic reasoning task was 19.2 ($SD = 4.2$). The mean number of consistent items answered correctly was 7.02 ($SD = 1.17$), the mean number of neutral items answered correctly was 6.71 ($SD = 1.58$), and the mean number of inconsistent items answered correctly was 5.49 ($SD = 2.05$). There were indications of a significant belief bias effect (i.e., difference from the neutral condition) on both the consistent items, $t(123) = 3.37$, $p < .01$, and the inconsistent items, $t(123) = 8.64$, $p < .001$, although the effect was larger for the latter. As an overall index of belief bias, we employed the difference score, which was obtained by subtracting the number correct on the inconsistent items from the number correct on the consistent items. This belief bias index could, in principle, range from -8 to $+8$, but the actual range was from -1 to 8. Positive scores indicate some degree of belief bias, with the higher scores indicating more belief bias. The mean belief bias score was 1.53 ($SD = 1.77$), and it was significantly different from zero, $t(123) = 9.62$, $p < .001$; 82 of 124 participants displayed some degree of belief bias.

Is Projection of Prior Belief in Height Judgment Associated With Belief Bias in a Verbal Task?

We now turn to the question of whether there is any domain generality to the projection of prior belief across verbal and nonverbal tasks. Several correlations in the matrix displayed in Table 1 are relevant to this question. The first two variables listed in Table 1 reflect the degree of projection of the gender relationship onto the estimates, r(G, E), in the matched and ecological sets, respectively. The degree of belief bias on the syllogistic reasoning task was significantly correlated with the former but not with the latter (see the fourth row of Table 1). The correlations were different in sign and they were significantly different from each other: .209 versus $-.095$, $t(114) = 3.52$, $p < .001$, test for difference in dependent correlations (see Cohen & Cohen, 1983, pp. 56–57). Thus, it is the condition where the participant is instructed to ignore the relationship between gender and height that correlates with belief bias in the syllogisms task and not the ecological condition where the gender relationship is in fact predictive.

We calculated another index from the height judgment task that reflects the ability to ignore prior knowledge and to follow the instructions in the matched set, which instructed the participants to not use gender as a height judgment cue. The need for this index derives from a property shared by both pure projection indices in the matched condition. The pure projection indices from the matched condition, $r(\text{G, E})_{\text{matched}}$ and $(\text{M} - \text{F})_{\text{matched}}$, should both be zero if the participant fully responded to the instructions and removed all influence of the prior belief—any deviation from the zero value is interpreted as an intrusion of prior knowledge into the height judgments. The problem comes about because reducing these indices to zero may be differentially difficult, not because of variation in the belief bias itself, but because the strength of the gender stereotype may vary from participant to participant. To put it simply, it should be easier for a participant who previously thought that there was on average, a 2-in. difference between males and females to reduce $(\text{M} - \text{F})_{\text{matched}}$ to zero than it would be for a participant who thought that there was a 7-in. difference between males and females. The prior belief of the latter participant is much more strongly contradicted by the matched set. Fortunately, in the ecological set, we actually have a measure of the participant's prior belief about the strength of the gender–height relationship, $(\text{M} - \text{F})_{\text{ecological}}$. We used this estimate of the prior belief to construct an index of the degree of adjustment from the ecological set that was achieved in the matched set. This index, HJ Adjustment (height judgment adjustment), was simply $(\text{M} - \text{F})_{\text{ecological}}$ minus $(\text{M} - \text{F})_{\text{matched}}$—the mean difference between male

Table 1
Intercorrelations Among the Primary Variables

Variable	1	2	3	4	5	6	7	8	9
Belief bias indices									
1. $r(G, E)_{matched}$	—								
2. $r(G, E)_{ecological}$	.526	—							
3. HJ Adjustment Score	−.396	.342	—						
4. Syllogisms, Belief Bias	.209	−.095	−.341	—					
Cognitive decontextualization tasks									
5. Covariation Judgment	−.200	.168	.322	−.342	—				
6. AET	−.261	.101	.284	−.236	.324	—			
Cognitive ability measures									
7. CA1	−.155	.256	.347	−.445	.493	.449	—		
8. CA2	−.215	.263	.461	−.495	.505	.392	.707	—	
Styles of epistemic regulation									
9. AOT	−.130	.057	.119	−.315	.241	.436	.358	.323	—

Note. Correlations larger than .182 are significant at the .05 level (two-tailed); correlations larger than .237 are significant at the .01 level (two-tailed); and correlations larger than .299 are significant at the .001 level (two-tailed). $r(G, E)_{matched}$ = correlation between gender and estimates in the matched condition; $r(G, E)_{ecological}$ = correlation between gender and estimates in the ecological condition; HJ Adjustment Score = $(M - F)_{ecological}$ minus $(M - F)_{matched}$; CA1 = cognitive ability composite 1 (WAIS block design and WAIS vocabulary); CA2 = cognitive ability composite 2 (Raven Matrices and vocabulary); AET = argument evaluation test; AOT = composite actively open-minded thinking scale.

and female height estimates in the ecological set minus the mean difference in the matched set. This index captures how much the participant succeeded in reducing his or her estimates of male–female difference (use of a proportional score rather than a difference score produced very similar results). The mean HJ Adjustment score was 2.92 ($SD = 1.2$)—the average participant decreased the male–female difference in their estimates by almost 3 in. The range in this index was quite large, from .03 in. to 7.41 in.

Presented in the third column of Table 1 are the correlations involving the HJ Adjustment index. Its correlation with belief bias on the syllogistic reasoning task was even larger in absolute magnitude than the simpler belief projection index, $r(G, E)_{matched}$ (−.341 vs. .209; the correlation with the former is negative because those showing larger adjustments to the matched set instructions displayed less belief bias in syllogistic reasoning).

Is Projection of Prior Belief Associated With Cognitive Ability and Cognitive Decontextualization?

The two composite indices of general cognitive ability, CA1 and CA2, were both negatively correlated with projection in the matched set and positively correlated with projection in the ecological set. In both cases, the correlations with $r(G, E)_{matched}$ and $r(G, E)_{ecological}$ were significantly different, $t(114) = 4.98$, $p < .001$, and $t(114) = 6.02$, $p < .001$, in the cases of CA1 and CA2, respectively. Consistent with these findings are the significant moderate correlations with HJ Adjustment displayed by CA1 and CA2 (.347 and .461, respectively).

Similar patterns were apparent for the two cognitive decontextualization tasks—the covariation judgment task and the AET. They were both significantly correlated with belief projection in the matched set but not with the belief projection in the ecological set. In both cases, the signs of the pairs of correlations were different, and each pair of correlations was significantly different from each other, $t(114) = 4.30$, $p < .001$, and $t(113) = 4.34$, $p < .001$, for the AET and covariation task, respectively. Significant correlations were observed between the HJ Adjustment score and the covariation judgment task (.322) and the HJ Adjustment score and the AET (.284).

In summary, cognitive ability and decontextualization skill was positively correlated with belief bias in a perceptual judgment task where prior belief accurately reflected an aspect of the perceptual environment and negatively correlated with belief bias in a perceptual judgment task where prior belief was incongruent with the perceptual environment. This finding is consistent with the notion of human intelligence as adaptive to context (Sternberg, 1985, 1997).

Given the findings on the height judgment task, it then becomes interesting to examine whether cognitive ability and decontextualization skill are associated with belief bias on a verbal reasoning task. Both indices of cognitive ability, CA1 and CA2, displayed significant and moderate negative correlations with belief bias on the syllogistic reasoning task (−.445 and −.495), as did performance on the covariation task (−.342) and the AET (−.236). The direction of all of these correlations was consistent: Individuals higher in cognitive ability displayed less belief bias, and they scored higher on the two measures of the ability to reason independently of prior belief (the AET and the covariation judgment task).

Relationships With a Composite Actively Open-Minded Thinking Score

Because the eight subscales on the actively open-minded thinking questionnaire displayed moderate intercorrelations

Table 2
Hierarchical Regression Analyses

Step	Variable	R	R^2_{change}	F_{change}	Final F	Partial r
			Predicting belief bias in syllogistic reasoning			
1	CA1	.445	.198	30.21**	19.86**	−.375
2	AOT	.476	.028	4.36*	4.36*	−.186
			Predicting argument evaluation test performance			
1	CA1	.449	.202	30.80**	16.76**	.349
2	AOT	.537	.086	14.77**	14.77**	.330
			Predicting covariation judgment performance			
1	CA1	.493	.243	38.81**	30.10**	.349
2	AOT	.497	.004	0.66	0.66	.074

Note. CA1 = cognitive ability composite 1 (WAIS block design and WAIS vocabulary); AOT = composite actively open-minded thinking scale.
$^*p < .05$. $^{**}p < .01$.

(see Appendix B, Table B1), a composite actively open-minded thinking (AOT) score was formed. First, the total scores on each of the subscales were standardized. Then, the standard scores on the flexible thinking, openness–ideas, and openness–values subscales were summed. From this sum was subtracted the sum of the standard scores on the absolutism, dogmatism, categorical thinking, belief identification, and need for closure subscales (the creation of a composite score using the factor score from the first factor produced results virtually identical to those from the unit-weighted sum of standard scores). 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 the composite score was justified by a factor analysis conducted on the eight subscales (see Appendix B, Table B2). Only one factor had an eigenvalue greater than one. All variables had loadings greater than .300 on this factor. This factor accounted for 38.7% of the variance, whereas the second factor accounted for only 6.1% of the variance.

The actively open-minded thinking composite score did not significantly correlate with projecting the gender relationship in the height judgment task, but it did display significant correlations with belief bias on the syllogistic reasoning task (see last line of Table 1). This result is consistent with previous demonstrations that measures of epistemic regulation can predict belief bias in verbal tasks (Kardash & Scholes, 1996; Klaczynski et al., 1997). In fact, the actively open-minded thinking composite score predicted belief bias in the syllogistic reasoning task even after differences in cognitive ability were partialled out. A multiple R of .476 was observed when the actively open-minded thinking composite score and the prorated WAIS IQ score (CA1; similar results were obtained with CA2) were entered as predictors of belief bias in syllogistic reasoning (see the first hierarchical multiple regression analysis in Table 2). As indicated in Table 2, CA1 was a significant unique predictor (partial correlation = −.375), but so was the actively open-minded thinking composite score (partial correlation = −.186). This trend was even more apparent in the next analysis in Table 2, where performance on the AET was the criterion variable. There, the actively open-minded thinking composite score accounted for 8.6% unique variance after CA1 was entered into the regression equation ($p < .01$), and its partial correlation was almost as large as that of the cognitive ability measure (.330 vs. .349). However, a different outcome was obtained when using covariation judgment performance as the criterion variable. There, despite a significant zero-order correlation (see Table 1), the AOT composite score did not account for significant variance after cognitive ability was entered into the regression equation.

General Discussion

In response to the relatively modest reduction in the gender–height stereotype that they achieved with their matched set manipulation, Nelson et al. (1990) commented that "people may be largely unable to control the influence of real-life base rates (e.g., the stimulus–response association between sex and height) that have been built up over a lifetime of experience, despite their best attempts to do so" (p. 672). We observed a somewhat more substantial reduction (from a mean height difference of 4.13 in. to one of 1.21 in.), but more importantly, we have documented very large individual differences in the ability to "control the influence of real-life base rates" and have shown that the ability to control the influence of prior knowledge in this perceptual judgment task is related to the ability to avoid belief bias on a verbal reasoning task. Scores on two different measures of cognitive ability (CA1 and CA2) and two different measures of cognitive decontextualization (AET and covariation judgment task) were also related to the ability to separate height judgments from the gender stereotype. Thus, people can "control the influence of real-life base rates," but there are large individual differences in the ability to do so—differences that are in part predictable from other reasoning and problem solving abilities.

That individual differences in the ability to avoid gender projection in the matched set is predictable is related to the issue of the domain specificity and generality raised in the introduction. An extreme form of domain specificity—one where the ability to evaluate evidence in an unbiased manner is completely independent from domain to domain—was clearly falsified by the results presented here. The ability to separate the influence of prior knowledge from the estimate required in the height judgment task (a perceptual judgment situation) was linked to the ability to separate prior knowledge from judgments of logical validity in the syllogistic reasoning task (a purely verbal task).

Although an extreme form of the domain specificity view is falsified by these results, the results regarding the actively open-minded style of epistemic regulation were somewhat more supportive. There, it was found that the relationships with a verbal reasoning measure of belief bias (the syllogistic reasoning task) were quite different from those involving belief bias in the height judgment task. In the latter, there were no significant zero-order correlations with the actively open-minded thinking composite score. However, in the former, not only was the zero-order correlation significant, but the actively open-minded thinking composite score accounted for variance in the belief bias effect even after the variance associated with a measure of cognitive ability had been partialled out.

The actively open-minded thinking (AOT) composite score was also significantly correlated with performance on both of the cognitive decontextualization tasks: the AET and the covariation judgment task. However, the AOT composite score was a unique predictor (after cognitive ability was partialled out) of the former but not the latter (see Table 2). The difference in the importance of epistemic regulation across these two tasks might relate to a distinction introduced by Wilson and Brekke (1994) in their classification of reasoning errors. They distinguished reasoning errors involving the failure to apply an appropriate rule from those occurring in situations where there exist no appropriate rule or algorithm that guarantees problem solution. The second situation is conducive to what Wilson and Brekke termed *mental contamination*—unwanted influences on judgments and evaluations. Whereas an incorrect evaluation of the data in the covariation judgment task reflects the failure to approximate a rule (the Δp rule of conditional probability), there is no clear-cut rule or algorithm to apply to the argument-evaluation items. Thus, the AET may be susceptible to mental contamination from prior beliefs in a manner not as obvious in the covariation judgment task. Epistemic regulation in the context of the AET might consist of metacognitive recognition of the lack of definitive response rules, followed by attempts to control the influence of the likely contaminating effects of prior belief. Our results from the AOT scale demonstrated that indirect indicators of tendencies toward such epistemic regulation predicted performance on the AET over and above the variance accounted for by cognitive ability.

The results involving the cognitive ability measures were very much in line with an adaptive interpretation of the nature of human intelligence (Larrick, Nisbett, & Morgan, 1993; Sternberg, 1985, 1997; Sternberg & Detterman, 1986). Both cognitive ability measures (CA1 and CA2) were positively correlated with belief bias in a perceptual judgment task where prior knowledge accurately reflected an aspect of the perceptual environment (the ecological set) and negatively correlated with belief bias in a perceptual judgment task where prior knowledge was incongruent with the perceptual environment (the matched set)—just the pattern that an adaptive view of intelligence would predict.

Our results on styles of epistemic regulation are consistent with those of Klaczynski et al. (1997), but our results with respect to cognitive ability are not. They found no correlation between cognitive ability and belief bias in evaluating arguments and evaluating evidence, whereas we found significant correlations on our syllogistic reasoning measure. Aside from some clear differences in the tasks, we think one difference might be that our participants, unlike theirs, were explicitly and specifically instructed to ignore prior knowledge and belief in all our tasks. Thus, our situations might have been more prone to implicate metacognitive control processes that are strongly identified with intelligence (Byrnes, 1995; Sternberg, 1985).

Cognitive ability was associated with skill at detaching prior knowledge from reasoning in the syllogistic reasoning task. If contextualist theories were correct, and tendencies toward decontextualization are actually maladaptive, then people with increased cognitive capacity might be expected to display greater reluctance to detach world knowledge from their information processing. Thus, a strong version of the contextualist position—that, because of the primacy of context-based processing (Hilton, 1995; Levinson, 1995; Stanovich, 1999), more cognitively able individuals would be more likely to carry over contextual information into an environment where context was inefficacious—was not supported. Our results concerning cognitive ability thus seem to bolster the normative status of styles of epistemic regulation that emphasize unbiased reasoning (styles that are focal to the critical thinking literature), but with an important caveat that is consistent with the contextualist tradition. More intelligent individuals do contextualize the problem more when that context contains cues that can facilitate judgment, but they are less likely to carry over contextual cues into situations where they know the cues are no longer diagnostic.

Generating at least contingent support for the normative tradition in the critical thinking literature is important because several philosophical analyses have called into question the normative status of the stricture that evidence evaluation not be contaminated by prior belief. For example, Kornblith (1993) argued:

> 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)

This argument—that in a natural ecology where most of our prior beliefs are true, projecting our beliefs onto new data will lead to faster accumulation of knowledge—has been termed the knowledge projection argument (see Stanovich, 1999), and it reappears in a remarkably diverse set of contexts throughout the reasoning and decision making literature (Dawes, 1990; Edwards & Smith, 1996; Evans et al., 1993; Koehler, 1993; Krueger & Zeiger, 1993; see Stanovich, 1999, for a review).

The contingent validity of this argument and its interaction with individual differences is perhaps best conceptualized within two-process models of cognition. Specifically, numerous theorists have proposed two-process models of cognitive activity that distinguish automatic heuristic processing from controlled analytic processing (e.g., Epstein, 1994; Evans & Over, 1996; Sloman, 1996; see Stanovich, 1999, for a review). Instructions requiring decontextualization (such as instructions to ignore a prior belief) probably heavily stress the analytic system, which must be employed to override the heuristic system when the latter's contextualized response is inappropriate (Stanovich, 1999). Because analytic processing is more closely related to computational capacity than is heuristic processing (Evans & Over, 1996; McGeorge, Crawford, & Kelly, 1997; Reber, 1993), under the two-process view it might be expected that successful decontextualization would be associated with higher cognitive ability.

References

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. *American Educational Research Journal, 31,* 313–337.

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15,* 147–149.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11.

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.

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

Baron, J. (1985b). What kinds of intelligence components are fundamental? In S. Chipman & J. Segal (Eds.), *Thinking and learning skills* (Vol. 2, pp. 365–390). Hillsdale, NJ: Erlbaum.

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). *Thinking and deciding (2nd ed.).* New York: Cambridge University Press.

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning, 1,* 221–235.

Baron, J. B., & Sternberg, R. (Eds.). (1987). *Teaching thinking skills: Theory and practice.* New York: Freeman.

Biek, M., Wood, W., & Chaiken, S. (1996). Working knowledge, cognitive processing, and attitudes: On the determinants of bias. *Personality and Social Psychology Bulletin, 22,* 547–556.

Broniarczyk, S., & Alba, J. W. (1994). Theory versus data in prediction and correlation tasks. *Organizational Behavior and Human Decision Processes, 57,* 117–139.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100,* 511–534.

Byrnes, J. P. (1995). Domain-specificity and the logic of using general ability as an independent variable or covariate. *Merrill-Palmer Quarterly, 41,* 1–24.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60,* 1239–1249.

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.

Ceci, S. J. (1993). Contextual trends in intellectual development. *Developmental Review, 13,* 403–435.

Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development (Expanded ed.).* Cambridge, MA: Harvard University Press.

Cederblom, J. (1989). Willingness to reason and the identification of the self. In E. P. Maimon, B. F. Nodine, & F. W. O'Conner (Eds.), *Thinking, reasoning, and writing* (pp. 147–159). New York: Longman.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory.* Odessa, FL: Psychological Assessment Resources.

Dawes, R. M. (1990). The potential nonfalsity of the false consensus effect. In R. M. Hogarth (Ed.), *Insights into decision making* (pp. 179–199). Chicago: University of Chicago Press.

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

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71,* 5–24.

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.

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.

Evans, J. S. B. T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11,* 295–306.

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

Evans, J. S. B. T., Over, D. E., & Manktelow, K. (1993). Reasoning, decision making and rationality. *Cognition, 49,* 165–187.

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

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

Kao, S. F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of participative cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1363–1386.

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.

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.

Klaczynski, P. A., Gordon, D. H., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology, 89,* 470–485.

Klaczynski, P. A., & Narasimham, G. (1998). Development of scientific reasoning biases: Cognitive versus ego-protective explanations. *Developmental Psychology, 34,* 175–187.

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.

Krueger, J., & Zeiger, J. (1993). Social categorization and the truly false consensus effect. *Journal of Personality and Social Psychology, 65,* 670–680.

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology, 65,* 861–876.

Kuhn, D. (1991). *The skills of argument.* Cambridge, England: Cambridge University Press.

Kuhn, D. (1996). Is good thinking scientific thinking? In D. R. Olson & N. Torrance (Eds.), *Modes of thought: Explorations in culture and cognition* (pp. 261–281). New York: Cambridge University Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498.

Larrick, R. P., Nisbett, R. E., & Morgan, J. N. (1993). Who uses the cost-benefit rules of choice? Implications for the normative status of microeconomic theory. *Organizational Behavior and Human Decision Processes, 56,* 331–347.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, England: Cambridge University Press.

Levin, I. P., Wasserman, E. A., & Kao, S. F. (1993). Multiple methods of examining biased information use in contingency judgments. *Organizational Behavior and Human Decision Processes, 55,* 228–250.

Levinson, S. C. (1995). Interactional biases in human thinking. In E. Goody (Ed.), *Social intelligence and interaction* (pp. 221–260). Cambridge, England: Cambridge University Press.

Lipman, M. (1991). *Thinking in education.* Cambridge, England: Cambridge University Press.

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.

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

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.

McGeorge, P., Crawford, J., & Kelly, S. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 239–245.

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.

Moshman, D., & Franks, B. (1986). Development of the concept of inferential validity. *Child Development, 57,* 153–165.

National Dairy Council. (1980). *Recommended dietary allowances revised* (NASNRC Publication No. ISSN 0111-5568). Rosemont, IL: Author.

Nelson, T., Biernat, M., & Manis, M. (1990). Everyday base rates (sex stereotypes): Potent and resilient. *Journal of Personality and Social Psychology, 59,* 664–675.

Nickerson, R. S. (1987). Why teach thinking? In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 27–40). New York: Freeman.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking.* Pacific Grove, CA: Midwest Publications.

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., 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.

Raven, J. C. (1962). *Advanced Progressive Matrices* (Set II). London: H. K. Lewis.

Raven, J. C., Court, J. H., & Raven, J. (1977). *Manual for Advanced Progressive Matrices* (Sets I & II). London: H. K. Lewis.

Reber, A. S. (1993). *Implicit learning and tacit knowledge.* New York: Oxford University 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.

Sattler, J. M. (1988). *The assessment of children* (3rd ed.). San Diego: Jerome Sattler Publishers.

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.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 101–120.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology, 48A,* 257–279.

Siegel, H. (1988). *Educating reason.* New York: Routledge.

Siegel, H. (1997). *Rationality redeemed? Further dialogues on an educational ideal.* New York: Routledge.

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

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.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition, 20,* 51–68.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89,* 342–357.

Stanovich, K. E., & West, R. F. (1998a). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127,* 161–188.

Stanovich, K. E., & West, R. F. (1998b). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition, 28,* 161–179.

Stanovich, K. E., West, R. F., & Harrison, M. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31,* 811–826.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge, England: Cambridge University Press.

Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist, 52,* 1030–1037.

Sternberg, R. J., & Detterman, D. K. (1986). *What is intelligence?* Norwood, NJ: Ablex.

Stigler, J. W., & Baranes, R. (1988). Culture and mathematics learning. In E. Z. Rothkopf (Ed.), *Review of Research in Education* (Vol. 15, pp. 253–306). Washington, DC: American Educational Research Association.

Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics.* New York: Harper & Row.

Troldahl, V., & Powell, F. (1965). A short-form dogmatism scale for use in field studies. *Social Forces, 44,* 211–215.

Wade, C., & Tavris, C. (1993). *Critical and creative thinking.* New York: HarperCollins.

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 509–521.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67,* 1049–1062.

Wechsler, D. (1981). *Manual for Wechsler Adult Intelligence Scale—Revised.* New York: Psychological Corporation.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). San Diego, CA: Academic Press.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

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.

(Appendixes follow)

Appendix A

Items From Two Previously Unpublished Subscales

Items on the Flexible Thinking Subscale

1. Changing your mind is a sign of weakness. (Reflect)
2. A person should always consider new possibilities.
3. Intuition is the best guide in making decisions. (R)
4. If I think longer about a problem I will be more likely to solve it.
5. Basically, I know everything I need to know about the important things in life. (R)
6. Considering too many different opinions often leads to bad decisions. (R)
7. People should always take into consideration evidence that goes against their beliefs.
8. Difficulties can usually be overcome by thinking about the problem, rather than through waiting for good fortune.
9. There is nothing wrong with being undecided about many issues.
10. Coming to decisions quickly is a sign of wisdom. (R)

Items on the Belief Identification Subscale

1. What beliefs you hold have more to do with your own personal character than the experiences that may have given rise to them.
2. It is a noble thing when someone holds the same beliefs as their parents.
3. One should disregard evidence that conflicts with your established beliefs.
4. Someone who attacks my beliefs is not insulting me personally. (R)
5. It is important to persevere in your beliefs even when evidence is brought to bear against them.
6. Certain beliefs are just too important to abandon no matter how good a case can be made against them.
7. Abandoning a previous belief is a sign of strong character. (R)
8. Beliefs should always be revised in response to new information or evidence. (R)
9. It makes me happy and proud when someone famous holds the same beliefs that I do.

Appendix B

Table B1
Intercorrelations Among the Subscales on the Epistemic Regulation Questionnaire

Subscale	1	2	3	4	5	6	7	8
1. Flexible Thinking	—							
2. Openness-Ideas	.250	—						
3. Openness-Values	.459	.171	—					
4. Absolutism	−.424	−.347	−.623	—				
5. Dogmatism	−.320	−.101	−.490	.513	—			
6. Categorical Thinking	−.254	−.060	−.456	.557	.550	—		
7. Belief Identification	−.527	−.245	−.523	.549	.378	.403	—	
8. Need for Closure	−.193	−.149	−.147	.356	.200	.167	.148	—

Note. Correlations larger than .176 are significant at the .05 level (two-tailed); correlations larger than .231 are significant at the .01 level (two-tailed); correlations larger than .293 are significant at the .001 level (two-tailed).

Table B2
Component Loadings for all Subscales After Factor Analysis and Varimax Rotation

	Component	
Subscale	1	2
Flexible Thinking	−.588	.305
Openness-Ideas	−.303	.312
Openness-Values	−.728	
Absolutism	.833	
Dogmatism	.651	
Categorical Thinking	.667	.424
Belief Identification	.690	
Need for Closure	.302	
% variance accounted for	38.7%	6.1%

Note. Component loadings lower than .300 have been eliminated.

Received October 20, 1998
Revision received December 15, 1998
Accepted February 4, 1999 ■