Prior Belief Influences on Reasoning and Judgment: A Multivariate Investigation of Individual Differences in Belief Bias

by

Walter Cabral Sá

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Education
University of Toronto

© Copyright by Walter C. Sá 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-41304-7
Abstract

Prior Belief Influences on Reasoning and Judgment: A Multivariate Investigation of Individual Differences in Belief Bias

Doctor of Philosophy, 1999
Walter Cabral Sá

Graduate Department of Education
University of Toronto

Belief bias occurs when reasoning or judgments are found to be overly influenced by prior belief at the expense of a normatively prescribed accommodation of all the relevant data. The normative status and domain specificity / generality of belief bias was investigated in a multivariate investigation. The research reported here also examined the role of cognitive capacity and thinking dispositions on belief bias. Cognitive capacity deals with those parameters of interest to traditional tests of psychometric intelligence whereas thinking dispositions captures those parameters that deal with what can be thought of as the goals of the reasoner--how the capacities are put to use. A total of 124 participants recruited through an undergraduate university population participated in the study. Results suggested that belief bias is best interpreted as a nonnormative tendency in human reasoning and judgment. Furthermore, some moderate correlations among indices of belief bias across domains as varied as deductive reasoning and perceptual judgment suggest belief bias has some degree of domain generality. Finally, the results also suggest that both cognitive capacity and thinking dispositions account for some unique variance in belief bias. The results for this last finding, however, was moderated by the type of reasoning or judgments participants were asked to make.
Acknowledgements

I thank my mother Maria, and my sisters Laura and Lesley for their personal support, encouragement, and love. Let me also thank here my father Albano, and my brother-in-law Robert. From my family I have learned to place great value on the truly important things in life.

To the students in the Stanovich lab, both past and present, I offer my thanks. My thinking on matters of relevance to this work has no doubt benefited from discussions with Cindy Kaizer, Carol Kelley, Jason Riis, Robin Sidhu, Ron Stringer, and Maggie Toplak. I would particularly like to thank Maggie for taking the time to read this work and offering several helpful comments and suggestions for its defense. For similar reasons, I also thank several other students and faculty at the Department of Human Development and Applied Psychology, too many to name, but particularly Keith Oatley.

To my committee members Geoffrey Fong, Chris Olsen, David Olson, and Uri Shafrir, a sincere thank you. Your engagement with this work has contributed enormously to my understanding and appreciation for the additional complexities involved in interpreting this body of data.

To my supervisor Keith Stanovich, I offer my sincerest thanks. Nothing I can confine to a few lines here will express my full gratitude for Keith’s extraordinary guidance, patience, and help. I simply could not have imagined a better mentor. Intellectually, professionally, and personally, I have greatly benefited from Keith’s efforts.

Finally, to my loving wife Junko, I’d like to offer a heartfelt thank you. It pains me to think of the enormous sacrifices you have made in the last few years. Your love and support is something I’ll always cherish. I dedicate this thesis to you.
# Table of Contents

**Introduction**

- Belief Bias and the Normative Issue ........................................... 2  
- Bayes Theorem as a Normative Model for Belief Revision ............... 3  
- Syllogistic Reasoning and the Normative Issue ............................... 10  
- Normative Models as Justified Benchmarks for Optimal Reasoning ... 15  
- Addressing the Status of Normative Models .................................. 18  
- Ecological Validity and the Projection of Belief ............................ 20  
- The Domain Specificity and Generality of Belief Bias .................. 22  
- Addressing the Domain Specificity and Generality Issue .............. 24  
- Predicting Belief Bias Using a Cognitive Capacity - Thinking Dispositions Framework ........................................... 27  
- The Research Strategy .............................................................. 34  
- Hypotheses .................................................................................. 35  

**Method**

- Participants ..................................................................................... 37  
- Prior Belief Reasoning and Judgment Tasks .................................. 38  
  - Syllogistic Reasoning Test .......................................................... 38  
  - Height Judgment Tasks .................................................................. 39  
  - Argument Evaluation Test ............................................................ 42  
  - Covariation Judgment Task ............................................................ 45  
  - Confounded Variable Task ............................................................. 49  
- Social Perception Tasks ................................................................ 52  
  - Opinion Prediction ........................................................................ 52  
  - Knowledge Prediction ..................................................................... 53  
- General Cognitive Ability Measures ............................................. 55  
- Thinking Dispositions Questionnaire ............................................ 57  
  - Flexible Thinking Scale ............................................................... 57  
  - Counterfactual Thinking Scale ....................................................... 58  
  - Absolutism .................................................................................... 58  
  - Dogmatism .................................................................................... 59  
  - Belief Identification ....................................................................... 59  
  - Impulsiveness ................................................................................ 60  
  - Deliberation ................................................................................... 60
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness-Values</td>
<td>60</td>
</tr>
<tr>
<td>Openness-Ideas</td>
<td>61</td>
</tr>
<tr>
<td>Categorical Thinking</td>
<td>61</td>
</tr>
<tr>
<td>Need for Closure</td>
<td>61</td>
</tr>
<tr>
<td>Faith in Intuition</td>
<td>61</td>
</tr>
<tr>
<td>Head Over Heart</td>
<td>62</td>
</tr>
<tr>
<td>Paranormal Beliefs</td>
<td>62</td>
</tr>
<tr>
<td>Superstitious Thinking</td>
<td>62</td>
</tr>
<tr>
<td>Consideration of Future Consequences Scale</td>
<td>62</td>
</tr>
<tr>
<td>Concept of Luck</td>
<td>63</td>
</tr>
<tr>
<td>Response Bias and Effort Checks</td>
<td>63</td>
</tr>
<tr>
<td>Impression Management</td>
<td>63</td>
</tr>
<tr>
<td>Effort Probe</td>
<td>63</td>
</tr>
<tr>
<td>Demographic and Print Exposure Measures</td>
<td>63</td>
</tr>
<tr>
<td>Procedure</td>
<td>64</td>
</tr>
<tr>
<td>Overview of the Methodological Approach</td>
<td>68</td>
</tr>
</tbody>
</table>

**Results and Discussion**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Differences in Prior Belief Task Performance</td>
<td>70</td>
</tr>
<tr>
<td>The Correlates of Normative Responses on the Prior Belief Tasks</td>
<td>72</td>
</tr>
<tr>
<td>Consequences of Projecting Belief on the Accuracy of Tracking the Environment</td>
<td>78</td>
</tr>
<tr>
<td>Exploring the Domain Specificity and Generality of Belief Bias</td>
<td>82</td>
</tr>
<tr>
<td>Does Cognitive Ability Alone Provide the Underlying Mechanism for the Observed Degree of Domain Generality?</td>
<td>87</td>
</tr>
<tr>
<td>The Role of Thinking Dispositions on Prior Belief Task Performance</td>
<td>95</td>
</tr>
<tr>
<td>Is the Association Between Participants Giving the Normative Response on the Prior Belief Tasks and the Thinking Disposition Measures a Product of Those Participants Giving Socially Desirable Responses?</td>
<td>100</td>
</tr>
<tr>
<td>Do Thinking Dispositions Account for any Additional Variance in the Prior Belief Tasks Once the Variance Accounted for by Cognitive Ability is Partialled Out?</td>
<td>102</td>
</tr>
<tr>
<td>How Much Shared Variance Exists Between Thinking Dispositions and Cognitive Ability Measures?</td>
<td>108</td>
</tr>
</tbody>
</table>
Further Explorations of Thinking Dispositions: Epistemic Development in Higher Education .................................................. 111
Performance on the Projection Tasks ........................................ 113
Individual Differences in Projection ........................................ 114

Summary and Conclusions .......................................................... 128
Belief Bias and the Normative Issue ........................................... 128
The Domain Specificity and Generality of Belief Bias ................. 131
The Utility of the Cognitive Capacity - Thinking Dispositions
Framework in the Prediction of Belief Bias ............................... 132
Limitations of the Present Study ................................................. 138
Conclusions .............................................................................. 140

References .................................................................................. 144
List of Tables

Table 1. Order of Task Administration.................................................................65
Table 2. Summary of Tasks in the Testing Battery That Were
Not Completed in Proper by Participants..........................................................67
Table 3. Mean Scores of Participants with the Lowest and
Highest Scores on the Normative Performance Index
for the Syllogisms Test, Argument Evaluation Test
Covariation Judgment Task, and Confounded Variable Task...........74
Table 4. Mean Scores of Participants with the Lowest and
Highest Scores on the Normative Performance Index
for the Syllogisms Test, Argument Evaluation Test, Covariation
Judgment Task, and Confounded Variable Task.............................77
Table 5. Intercorrelations Among the Primary
Prior Belief Task Variables...........................................................................84
Table 6. Hierarchical Regression Analyses Predicting Projection
of Prior Belief on the Matched Height Judgment Task.........................86
Table 7. Hierarchical Regression Analyses Predicting Projection
of Prior Belief on the Matched Height Judgment Task with
Cognitive Capacity Measures Entered First.........................................89
Table 8. Hierarchical Regression Analysis Predicting Belief Bias
Score on Syllogism Test with Cognitive Capacity Measures
Entered First..............................................................................................91
Table 9. Hierarchical Regression Analysis Predicting Total Syllogism
Test Score with Cognitive Capacity Measures Entered First...............92
Table 10. Hierarchical Regression Analyses Predicting Syllogism
Problem-Type Performance with Cognitive Capacity Measures
Forced in the Model First.............................................................................93
Table 11. Hierarchical Regression Analyses Predicting Beta Weight for
Argument Quality on the Argument Evaluation Test with
Cognitive Capacity Measures Entered First..........................................94
Table 12. Hierarchical Regression Analyses Predicting Beta Weight
for the Delta p Statistic in the Covariation Judgment Task With
Cognitive Capacity Measures Entered First...........................................95

-vii-
Table 13. Intercorrelations Among the Primary Thinking Disposition Variables

Table 14. Component Loadings for Thinking Disposition Variables after Varimax Rotation

Table 15. Zero-Order Correlations Between Various Indices on the Prior Belief Tasks with the Composite Thinking Disposition Scales for Actively Open-Minded Thinking (AOT), Conscientiousness (Cons), Superstitious Thinking (ST) and the Impression Management Scale

Table 16. Hierarchical Regression Analyses Predicting Beta Weight for Argument Quality on the AET with Cognitive Capacity Measures Entered First Followed by Thinking Disposition Scales

Table 17. Hierarchical Regression Analyses Predicting Various Performance Indices on the Syllogistic Reasoning Test with Cognitive Capacity Measures Entered First followed by Thinking Disposition Scales

Table 18. Hierarchical Regression Analyses Predicting Performance Indices on the Covariation Judgment Task, Confounded Variable Task and Height Judgment Task with Cognitive Capacity Measures Entered First followed by Thinking Disposition Scales

Table 19. Component Loadings on a Two Factor Model for all Actively Open-minded Thinking Disposition Subscales, Cognitive Ability measures and the AET Beta for Argument Quality after Varimax Rotation

Table 20. Commonality Analyses of Thinking Disposition Measures and Cognitive Ability Measures for all Prior Belief Tasks

Table 21. Mean Percentage Estimate of Endorsement of Each Item on the Opinion Prediction Task by Subjects who Endorsed and Who Did Not Endorse the Item

Table 22. Mean Percentage Estimate of Correct Response for Each Item on the Knowledge Prediction Task by Subjects Who Incorrectly Answered and Correctly Answered the Item
Table 23. Zero-Order Correlations Between the r(T,P) Predictive Accuracy Indices of the Projection Tasks with Performance Indices of the Prior Belief Tasks, Cognitive Ability Measures and the Actively Open-Minded Thinking Scale (AOT) Composite

Table 24. Hierarchical Regression Analyses Predicting Predictive Accuracy on the Projection Tasks with Cognitive Capacity Measures Entered First Followed by The Actively Open-Minded Thinking Scale (AOT)

Table 25. Commonality Analyses of Thinking Disposition Measures and Cognitive Ability Measures on the Predictive Accuracy Indices of The Opinion Prediction and Knowledge Prediction Tasks
List of Figures

Figure 1. Illustration of the four resulting cells of numerical data conforming to a 2x2 contingency table.................................47

Figure 2. Visual schematic of methodological approach and relationships of interest.................................................................69

Figure 3. Frequency distributions for some focal indices on the prior belief tasks......................................................................71

Figure 4. Path diagram predicting the aggregate height judgments for both the ecological and matched tasks..........................81
List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>Syllogism Examples</td>
<td>162</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Examples of Stimuli Used in Height Judgment Tasks</td>
<td>163</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Examples of AET Items</td>
<td>165</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Examples of Covariation Judgment Items</td>
<td>167</td>
</tr>
<tr>
<td>Appendix E</td>
<td>Confounded Variable Task</td>
<td>169</td>
</tr>
<tr>
<td>Appendix F</td>
<td>Vocabulary Checklist</td>
<td>172</td>
</tr>
<tr>
<td>Appendix G</td>
<td>Print Exposure Measures</td>
<td>173</td>
</tr>
</tbody>
</table>
A term coined in the deductive-logic reasoning literature, belief bias was initially applied as a descriptive label for the robust phenomenon which arises when there exists conflict between a deductive argument's logical validity and the empirical truth value of the argument's conclusion. People faced with these deductive problems seem to base their evaluation of the logical validity on the believability of the conclusion rather than properly tracking the logical syntax of the argument structure (Evans, 1989, Johnson-Laird, 1984; Markovits & Bouffard-Bouchard, 1992; Markovits & Nantel, 1989; Newstead, Pollard, Evans, & Allen, 1992; Oakhill & Johnson-Laird, 1985; Oakhill, Johnson-Laird & Garnham, 1989; Revlin, Leirer, Yopp, & Yopp, 1980). In other words, prior beliefs (viz., knowledge of the world) are exerting a substantial influence on reasoning despite the fact that these beliefs are irrelevant to the task as defined. Similar results from a wide array of research paradigms also converge with the suggestion that peoples' reasoning and judgments are systematically biased by prior beliefs. Prior belief dominance in reasoning and judgment is observed in experimental paradigms demonstrating the selective processing of belief-congruent and belief incongruent evidence when evaluating claims (e.g., Edwards and Smith, 1996; Klaczynski, 1997; Klaczynski & Gordon, 1996; Klaczynski, Gordon, & Fauth, 1997; Lord, Ross & Lepper, 1979), perceiving illusory correlations that are congruent with prior beliefs (e.g., Broniarczyk & Alba, 1994; Chapman & Chapman, 1967, 1969), persevering in an experimentally induced belief even after totally discrediting the data used to induce the belief in the first place (e.g., Anderson, 1983; Anderson, Lepper & Ross, 1980; Anderson & Sechler, 1986), failure to search for arguments favouring the opposing side of an issue—viz., “myside bias” (e.g., Baron, 1995; Perkins, 1989; see also Lord, Lepper & Preston, 1984), differentially interpreting
identical ambiguous behaviours enacted by members of different ethnic/racial backgrounds in accord with ethnic or racial stereotypes (e.g., Darley & Gross, 1983; Sagar & Schofield, 1980), an inability to ignore the gender cue when judging the height of photographed males and females (participants' being well informed of a manipulation which effectively eliminated the usual and expected diagnosticity of the gender cue; Nelson, Biernat & Manis, 1990), a failure to even differentiate evidence from prior beliefs (in this case a theory; Kuhn, 1991, 1992), the tendency to evaluate arguments congruent with prior beliefs as high in quality regardless of the objective quality, and vice versa for arguments which conflict with prior beliefs (Edwards & Smith, 1996; Stanovich & West, 1997, 1998a). There remains little dispute that the belief bias effects are both real and pervasive (Baron, 1985a; Evans, Barston & Pollard, 1983; Evans, 1989; Edwards & Smith, 1996; Stanovich & West, 1997).

The present research centers upon this general tendency of human reasoning and judgment to be overly influenced by prior beliefs, viz., belief bias. This research can be organized around three general themes of investigation: 1) the normative status of belief bias, 2) the domain specificity and generality of belief bias and 3) the prediction of belief bias using a cognitive capacity thinking dispositions framework. Although these three themes serve as a useful framework for organizational purposes, it will become clear that they have some degree of overlap and should not be thought of as completely distinct issues. All three themes will be investigated by administering the same participants several prior belief tasks and examining both the patterns of variability and covariance across these tasks as well as the emerging individual differences and their covariates.

Belief Bias and the Normative Issue

The term “belief bias” can be used either with or without normative
connotations. Without the normative connotations, the usage of “belief bias” is constricted to a purely descriptive label for the tendency of human reasoning and judgment to be strongly influenced by prior belief. As such, no value judgment is made on either the efficacy or detriment such a bias holds for human cognition. Numerous investigations of belief bias effects have gone beyond this mere descriptive use and have attributed nonnormative connotations (Evans et al., 1983; Evans, Newstead, Allen, & Pollard, 1994; George, 1995; Klaczynski & Gordon, 1996; Klaczynski, Gordon, & Fauth, 1997; Markovits & Nantel, 1989; Newstead, Pollard, Evans & Allen, 1992; Oakhill & Johnson-Laird, 1985; Oakhill, Johnson-Laird & Garnham, 1989). A common methodology of comparing the consequences of belief biased reasoning and judgment to the dictates of the relevant normative models—traditionally taken from logic, mathematics and decision science—has resulted in demonstrations of reasoning and judgments that systematically deviate from the optimal response outlined by the normative model. Insofar as the normative models are held to outline the optimal response, this methodology has thus yielded results supporting the nonnormative status of belief bias.

Recently, however, a barrage of criticisms have taken aim at the appropriateness of various normative models as the optimal standards of rational thinking. Of particular relevance to this research, are claims that some degree of belief bias might be normatively appropriate (e.g., Koehler, 1993). Because the status of the normative models plays a defining role in how we conceptualize belief bias, the normative status of belief bias can be directly addressed by an evaluation of the normative status of the relevant normative models—an issue that will be addressed later.

Bayes Theorem as a Normative Model for Belief Revision

Bayes theorem has been used as a normative model for how the confidence in
ones beliefs ought to be updated upon receipt of new information. Although there are numerous ways one can outline Bayes theorem, the most useful formulation for the present purposes is to express it in its odds form:

\[
\frac{P(H|D)}{P(\neg H|D)} = \frac{P(D|H)}{P(D|\neg H)} \times \frac{P(H)}{P(\neg H)}
\]

A discussion of Bayes theorem begins with defining a few parameters. The focal hypothesis (or belief) being entertained is denoted in the theorem by \( H \), the data under consideration is denoted by \( D \), and \( P \) refers to the probability. The theorem can be broken down into the three essential components. Numerically, the right-most segment of the formula outlines the probability of the focal hypothesis being true divided by the probability of the focal hypothesis not (-) being true. This segment of the formula represents the prior odds or the probability that the focal hypothesis is true prior to having received the new data. Skipping over to the formula component on the left, this segment refers to the posterior odds or the probability that the focal hypothesis is true after obtaining the new datum and amalgamating it with all else that is known. The middle segment of the formula is of particular interest to the present discussion. It is in this component of the theorem that belief bias exerts its influence.

The middle component of the above theorem is the likelihood ratio (or the diagnostic ratio). This component represents the diagnosticity or informative value of the data in reference to evaluating \( H \). Mathematically, the likelihood ratio represents the probability of the new datum being obtained, given that the focal hypothesis (\( H \)) is true, divided by the probability of the datum obtaining given that \( H \) is not true (\( \neg H \)). As an illustration of this point, consider the situation in which the probability of having obtained the received datum given the focal hypothesis is 0.95, i.e., \( P(D|H) = 0.95 \). At first glance a 0.95 probability of having obtained the received data, given the focal hypothesis, may strike one as
impressive. This large probability, however, is meaningless without also accounting for the probability that the data would have obtained given the that focal hypothesis is false, i.e., \(P(D \mid \neg H)\). Thus \(P(D \mid \neg H)\) is a crucial piece of information for assessing the likelihood ratio. As an example of this point, contrast the case in which the \(P(D \mid \neg H)\) component is 0.95 with the case in which the \(P(D \mid \neg H)\) component is 0.10. In the first case, \(P(D \mid H)/P(D \mid \neg H)\) i.e., the likelihood ratio is equal to 1 (0.95/0.95 = 1). Likelihood ratio values of 1 (or close to 1) indicate nondiagnostic data. The presence of the given datum does not in any way differentiate the competing hypotheses— in this case the datum was equally likely under the the condition where the focal hypothesis is true \((H)\) and the condition in which the focal hypothesis is false \((\neg H)\). This nondiagnostic datum results in a likelihood ratio of 1 which when multiplied by the the prior odds (as dictated by the theorem) simply results in the posterior odds being equal to the prior odds. The confidence one places on a held belief is not changed in any way by the receipt of new information.

The situation is different when \(P(D \mid \neg H) = 0.10\). In this case the presence of the datum does differentiate the condition in which the focal hypothesis is true \((H)\) and the condition in which the focal hypothesis is not true \((\neg H)\). The probability that the datum was obtained given \(H\) is much greater than given \(\neg H\) (0.95 > 0.10). This results in a likelihood ratio which deviates from the value of 1 and thus when multiplied out in the formula results in a modification from the prior to the posterior.

Bayes theorem is not proposed as a descriptive account of what humans are doing when engaged in belief revision. Rather, as a suggested normative model of belief revision, the role of Bayes theorem is to dictate what an unlimited and idealized processor of information ought to do if the goal was to optimize the process of belief revision. Normative models do not take into account limitations
of human information processing—nor are they constrained by psychological plausibility.

An example of a major deviation of human judgment from the dictates of the Bayesian model is the apparent neglect of base rate information, viz., the prior odds segment of the formula. Base rate neglect has been an intensely explored bias in human probabilistic reasoning (Bar-Hillel, 1980, 1990; Cosmides & Tooby, 1996; Doherty & Mynatt, 1990; Koehler, 1996; Stanovich & West, 1998b). Several experimental setups have been used to demonstrate this insensitivity to base rate information. A classic example of this sort of demonstration is Cassecells, Schoenberger & Graboys, 1978 much used medical diagnosis problem. Participants are asked to answer the following problem:

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person’s symptoms or signs?

Only about 18% of participants will respond with the normatively correct answer of 2%. The modal response given by just under half the participants is the estimate of 95% (see Cosmides & Tooby, 1996). This modal response is in clear neglect of the base rate information that only 1/1000 people have the disease. The shear magnitude of the deviation can be seen by just considering that the 5% probability of a false positive result is significantly larger than the 1/1000 probability of actually having the disease itself. In other words, even in the case where the test indicates the presence of the disease, it is still more probative that the disease is actually not present (due to a false positive) than that the disease is actually present. An underlying appreciation for this fact alone should signal that the modal 95% estimate of having the disease after a positive test is completely off the mark.
Interestingly, when the prior probabilities take the form of prior beliefs held by an individual, not only are the priors likely not to be neglected, they appear to actually dominate the data. A Bayesian analysis does not, of course, dictate that the priors should not be factored into an analysis. Prior belief is represented in the above equation by the prior odds component. It is in this prior odds portion of the equation that prior belief may exert its normatively due influence in this Bayesian model. The problem seems to be that in addition to exerting an influence on the base rate segment of the equation, humans also allow the prior belief component to creep into the assessment of the likelihood ratio as well. The influence of prior beliefs on the assessment of the diagnostic ratio violates an important Bayesian implicature, viz., that the evaluation of evidence should not be influenced by prior beliefs. In short, prior beliefs get factored not once as normatively required, but twice into the equation.

This phenomenon is nicely demonstrated in the seminal study by Lord, Ross and Lepper (1979). Lord et al., presented their participants with research study data on the effectiveness of capital punishment as a deterrent to murder. Participants had been selected via a screening process such that they either were clear proponents or opponents of capital punishment. The data these participants were given was mixed in such a manner as to have half of the informational content supporting capital punishment as a deterrent to murder and the other half suggesting that capital punishment was not a deterrent to murder. This information supplied to participants was in the form of research study data which included the study results, procedural details, critiques of the studies, and rebuttals to the critiques.

The Bayesian model makes some clear predictions of what should transpire in such a situation if people are engaging in normative reasoning. The proponents and opponents of capital punishment are presented with an identical
body of data. The crucial difference between these two groups is that they hold different priors. What the participants are engaged with upon receipt of this mixed data, in terms of the Bayesian model, is the assessment of the likelihood ratio. New data is in receipt and the current task is to determine the probability of that datum having obtained given their focal hypothesis, i.e., \( P(D|H) \) in relation to the probability of having obtained that datum given that their focal hypothesis is incorrect \( P(D|\neg H) \).¹ In brief, participants are evaluating the diagnosticity of the data, i.e., the likelihood ratio of the Bayesian model.

In the Bayesian formulation of this process, there is no role for the prior odds (prior belief) to influence the likelihood ratio. Thus although the two groups in the Lord et al., study should be expected to differ markedly on the prior odds component of the Bayesian theorem, they should resemble one another in the evaluation of the evidence (i.e., the likelihood ratio component)--they are after all in receipt of identical information! The natural consequence of this formulation of the problem is that the posterior odds for the two groups should be closer to each other than were their prior odds. Specifically, the valence of the evidence is held as a constant and thus should initiate some convergence between the two opposing view points. This is not, however, what is found to happen.

Not only did the posterior odds of the two groups not converge, the two groups actually polarized even further. After examining the mixed research data, both of Lord et al.'s groups reported even more extreme views than they had originally stated prior to the presentation of the mixed data. The results were not only normatively deviant but somewhat disturbing in their implications. As the authors of the study point out, if people holding opposing points of view on some issue nevertheless manage to find further support for their own mutually exclusive positions, then much of social science research, in which the nature of

---

¹ It has been demonstrated that even this is not always done. Participants often ignore the relevance of the \( P(D|\neg H) \) component and focus almost exclusively on \( P(D|H) \). Often termed Pseudodiagnosticity, I will gloss over this further issue as it is not central to the point at hand.
the data is often mixed or inclusive, will frequently "fuel rather than calm the fires of debate" (p. 2108).

Why did Lord et al.'s groups diverge even further upon being given an identical body of empirical data? Consultation of the Bayesian model is not necessary to see that something counter intuitive seems to be at work. An identical body of data should suggest the same information to those in its receipt. It should not, however, be taken as further support by both camps of reasoners holding mutually exclusive positions. Lord et al., explain their results using the terms "biased assimilation of evidence". Biased assimilation of evidence refers to the tendency for participants to accept confirming evidence at face value and to engage in more careful scrutiny of disconfirming evidence. What is absolutely crucial for this point is that this asymmetrical scrutiny of evidence is not initiated by some consideration of issues related to methodology or procedures, but rather, on the basis of the congruency of a supported conclusion with a prior belief, viz., "does the conclusion of the study agree with what I believe to be true?"

Applying the vocabulary of a Bayesian analysis, the apparent cause of this polarization in beliefs lies with the biased assessment of the likelihood ratio. Contrary to the dictates of Bayes theorem, likelihood ratio's are influenced by the priors. Components of the mixed data set were systematically weighted differently by the two groups. Data supporting conclusions congruent with initial priors tend to be taken at face value whereas data supporting conclusions incongruent with initial priors are scrutinized with the usual result of being deemed relatively "unimpressive" or poor in quality. An important Bayesian implicature is thus violated by this biased assimilation of evidence: Evaluation of evidence should not be influenced by the prior odds viz., prior belief. Evidence

\[^{2}\text{Edwards and Smith (1996) have obtained reading time data which is consistent with this view. In their study, Edwards and Smith found that participants took significantly longer to read arguments that they disagreed with as opposed to those arguments they agreed with. In a similar vein to Lord et al., they concluded that their participants were submitting belief incongruent arguments to a higher standard of scrutiny than they had to the belief congruent arguments.}\]
should be evaluated on the basis of its own merit, thus making issues such as appropriateness of study design, adequacy of measures etc. the only sort of relevant dimensions of influence. Prior beliefs are not relevant to judgments about these “structural” or purely “syntactic” issues inherent to this stage of data evaluation.

To summarize the main point, although the normative Bayesian model renders prior belief as irrelevant in the assessment of the structural or syntactic issues involved in data evaluation, it is clear from the descriptive account that prior beliefs do exert appreciable influence on the data evaluation stage. This biased evaluation of evidence quality can be seen in numerous experimental research studies and has been associated with various terminology such as biased assimilation (Lord et al., 1979), selective scrutiny (see Newstead, Pollard, Evans & Allen, 1992), disconfirmation bias (Edwards & Smith, 1996), goal-oriented biases (Klaczynski, Gordon, & Fauth, 1997). As Klaczynski et al. (1997) point out, much of what is common among these explanations is at least an implicit notion of depth of processing being moderated by the congruency between evidence and prior belief. Specifically, belief-enhancing information seems to be processed at a relatively shallow level whereas belief-threatening evidence is subjected to more extensive analysis and processed at a deeper level (see also, Denes-Raj & Epstein, 1994; Kunda, 1990; Liberman & Chaiken, 1992; Schaller, 1992). An extreme example of having prior beliefs intrude on evaluations that should be confined to purely structural or syntactical concerns can be found in the syllogistic reasoning literature.

Syllogistic Reasoning
and the Normative Issue

Syllogistic problems have been extensively used in research on human deductive reasoning. The subclass of syllogisms of interest to this research are those that are designed to assess performance error associated with the semantic
content of a syllogism. These content-loaded syllogisms can be described as having two dimensions. The first is that of logical syntax. This dimension concerns the logical form of the syllogism which either results in the syllogism being logically valid or invalid. The following syllogism is an example of a syllogistic argument having only the single dimension of logical syntax:

premise 1: All X are Y
premise 2: All Y are Z
conclusion: All X are Z

The above syllogism is devoid of any semantic content. It is an abstract argument with elements signified by the variables X, Y and Z which in themselves have no referential relation to real world knowledge. A second dimension can be introduced when real world content is injected into the syllogistic argument. The following syllogistic argument is an example of a syllogism having both the dimensions of logical syntax and empirical truth:

premise 1: All fish can swim
premise 2: tuna are fish
conclusion: tuna can swim

With the introduction of semantic content into the syllogistic argument, the reasoner’s beliefs about the world now become relevant to the argument. In the example above, the syllogism’s conclusion “tuna can swim” is a proposition which can be verified as empirically true of the world. This of course need not have been the case, as the syllogism may also incorporate a conclusion that is not true of the world. Thus two different cognitive operations can be performed on such a syllogism. First, the reasoner may focus on the logical syntax dimension and evaluate the validity of the syllogism, i.e., does the conclusion follow logically from the premises. Second, the reasoner may simply evaluate whether the content of the syllogistic argument is empirically true, i.e., is the conclusion
actually true in the real world. Performing the former, however, proves to be
difficult in the presence of having the empirical cue available. As evidenced by
performance on the logical evaluation task, people find it difficult to ignore the
empirical dimension and focus exclusively on the structural logic of the
syllogism. A participant who is asked to evaluate the logical validity of a
syllogism independently of whether the conclusion is true or untrue (viz., a
purely structural assessment) is in effect asked to set aside prior beliefs about the
world. Thus the semantic content of the syllogistic argument cues an empirically
based response which people find difficult to avoid when making a response that
should be based purely on syntactical or structural considerations.

As an example of the difficulty that arises in syllogistic reasoning, evaluate
the logical validity of the syllogism below keeping in mind that the truth of the
conclusion is irrelevant to the question of logical validity:

  premise 1: All flowers have petals
  premise 2: roses have petals
  conclusion: roses are flowers

Roses are indeed flowers. This is an empirical issue which in this case is true.
The above syllogistic argument, however, is at the same time logically invalid. To
help make this distinction more clear evaluate the next syllogism:

  premise 1: All X have Y
  premise 2: Z have Y
  conclusion: Z are X

The fact that both X and Z share the property of having Y, does not necessarily
make X and Z the same thing. They may indeed be the same thing, but there
simply exists insufficient information to make that strong claim. The above
syllogism has the identical logical form of the previous flowers/roses syllogism.
The only difference between these two syllogisms is that the real world content
Belief Bias

has been stripped away from the latter syllogism. Since the real world content in the flowers/roses syllogism is in conflict with the logical validity, this stripping away of the content has the effect of making the structurally identical syllogism easier to evaluate.

This crucial point is even more clearly illustrated in the following syllogism:

premise 1: All cars have windows
premise 2: houses have windows
conclusion: houses are cars

In the above syllogism, real world content is once again injected into the same structural argument. The crucial difference here is that rather than conflicting with the logical validity of the argument, the empirical dimension is constructed to be congruent with the logical dimension. Assessing the logical validity of this argument seems effortless. It seems clear that the conclusion does not follow logically from the two premises. The logical structure of this syllogism is once again identical to the flowers/roses syllogism above. They are both logically invalid, but unlike the cars/houses syllogism in which there exists a congruence between the empirical and logical dimensions, the incongruency between the two dimensions of the flowers/petals syllogism complicates the structural judgment by pitting prior belief against deductive logic reasoning. Note that the tracking of syntactic structure in the evaluation of the cars/houses syllogism is not facilitated by any greater reliance on deductive logic, but rather the facilitation of accurate performance is a consequence of the empirical evaluation converging with a logical evaluation. To summarize, when required to evaluate the logical validity of syllogistic arguments, much of the time people behave as if the task was to evaluate “whether the conclusion agrees with what is empirically true of the world” as opposed to “whether the conclusion follows logically from the premises”.
This bias in human logical reasoning with syllogistic arguments is where the term belief bias was first employed. Because this literature sets much of the framework for the discussion of belief bias in this paper, it would seem prudent to outline alternative interpretations for the apparent “belief bias” in syllogistic reasoning. All these alternative interpretations to the standard belief bias view have been systematically shown to be inadequate.

One such alternative explanation had researchers argue that the observed belief bias effect in syllogistic reasoning was attributable to conversion of the premises during their initial encoding (Revlin & Leirer, 1978; Revlin, Leirer, Yopp & Yopp, 1980). The essential argument was that when a reasoner is given the premise “All A are B” they may also erroneously encode this proposition to imply the biconditional “All B are A”—an interpretation not logically warranted by the original premise. Conversion effects were argued to account for much of the observed belief bias in the previous research which had unfortunately confounded syllogisms employing premises prone to biconditional interpretations with belief-incongruent syllogisms. Thus it was no longer clear which of these two potential factors led to people exhibiting what had been labeled as a belief bias effect. Evans, Barston, & Pollard (1983), however, effectively demonstrated that there were consistently large and significant belief bias effects despite the use of stringent controls for conversion of premises. What Evans et al., (1983) cleverly demonstrated was that belief bias effects were still evident even when the logical validity of the syllogisms used were not moderated by biconditional interpretations.

Another class of alternative accounts suggested that people were simply unclear about the logical necessity requirement of the task. The concern is that demonstrations of human failure are not the result of inherent biases of logical reasoning per se but rather a consequence of misinterpreting or not
understanding what the task requires them to do—as put by Henle (1962) participants “fail to accept the logical task”. Evans et al. (1983, see also Evans, Newstead, Allen & Pollard, 1994) demonstrated, however, that belief bias was found to persist despite the use of explicit instructions outlining the requirements of the syllogistic reasoning task and the meaning of logical necessity.

Yet another alternative explanation made in a similar vein to that above suggests participants had skipped the premises altogether and judged the validity of the syllogism on the basis of the conclusion’s empirical truth value alone. Although consideration of the research findings above also seem to place this criticism somewhere off the mark, a different methodological strategy shows this alternative account to be clearly untenable. Participants also displayed belief bias effects when they were required to generate their own conclusions to incomplete syllogisms consisting only of the premises (viz., a production task; Oakhill & Johnson-Laird, 1985; Markovitz & Nantel, 1989). Participants were once again prone to producing conclusions that were consistent with their beliefs regardless of their logical validity. The effect was so large, in fact, that participants in one study exhibited more belief bias in the production task than the traditional evaluation task (Markovitz & Nantel, 1989).

Normative Models as Justified Benchmarks for Optimal Reasoning

The ubiquitous use of syllogistic tasks in investigations of human reasoning is not only the consequence of the reliable and systematic nature of the errors people make on these logical reasoning problems. The generous use of syllogistic problems in reasoning research is also due to one other very nice aspect of the syllogism: There is consensus regarding the appropriate normative model. Lacking a clear consensus on what counts as correct and incorrect on a class of reasoning problem is one problem much of reasoning research has been plagued with. In at least this one regard, syllogisms are unique.
Consensus on a normative model, however, is not enough to hold back a critique about the relevance of using syllogisms as tools for investigating human reasoning. Concerns have been raised that deductive reasoning problems such as the syllogism are too far removed from the nature of “true” human reasoning. One such critic is J. R. Anderson who asserts “Deductive reasoning enables one to go from certain premises to certain conclusions ... certain premises are rare or nonexistent in situations of adaptive importance. As a consequence, there is no reason why humans should have evolved to engage in correct deductive reasoning” (Anderson, 1990, p. 37). In a typical Johnson-Laird fashion3, another way to express this argument is to restate the argument in the following manner:

premise 1: Deductive reasoning enables one to go from certain premises to certain conclusions

premise 2: Certain premises are rare or nonexistent in situations of adaptive importance

conclusion: There is no reason why humans should have evolved to engage in correct deductive reasoning

It would seem that at least at some level the ability to engage in syllogistic reasoning has some value. Setting aside this issue of whether or not the human cognitive apparatus should have adapted to the dictates of deductive logic, at minimum, it is clear that the modal response given by people does deviate in a systematic manner from what is unanimously regarded as the normatively correct response. The nature of the errors people make on the content loaded syllogisms is of particular interest to the present research. Thus syllogistic reasoning provides the quintessential example of the general inability to inoculate one against the influence of prior beliefs when it would be efficacious to do so.

3 Johnson-Laird points out that many critics’ arguments against the relevance of syllogistic reasoning in relation to real word reasoning oddly enough take the form of a syllogistic argument (Johnson-Laird, 1983, p. 71).
Whereas the status of the normative model used for syllogistic reasoning is not in dispute, Bayes theorem as a normative model for belief revision represents a more representative case in this field in which the normative model is very much under dispute. The central Bayesian implicature that prior belief should not influence the evaluation of evidence is a topic of hot contention. On observing the selective scrutiny of verbal arguments by their participants, Edwards and Smith (1996) suggested that such a “disconfirmation bias” can be normatively justified. The congruency between an argument and prior belief can be taken as a cue that a fallacy in the argument is unlikely. On the other hand, incongruency between argument and belief may be taken as a cue indicating a fallacy is present. Since scrutinizing arguments comes with some mentally costly processing, it is justifiable to scrutinize arguments selectively on the basis of this belief/argument congruency (Edward & Smith, 1996). Koehler (1993) had also raised this same concern with a further demonstration that the evaluation of evidence by professional scientists themselves is influenced by prior belief. Upon observing “biased assimilation” of evidence by their participants, Lord et al. (1979) had anticipated these arguments: “...the physicist would be “biased,” but appropriately so, if a new procedure for evaluating the speed of light were accepted if it gave the “right answer” but rejected if it gave the “wrong answer.” The same bias leads most of us to be skeptical about reports of miraculous virgin births or herbal cures for cancer, and despite the risk that such theory-based and experience-based skepticism may render us unable to recognize a miraculous event when it occurs, overall we are surely well served by our bias” (p. 2106).

Although acknowledging that there may be some justification to believe that evidence supporting one’s prior beliefs is of better quality than evidence that contradicts a prior belief, Lord et al. (1979) are quick to point out that the observed

---

4 Edwards and Smith (1996), however, make it clear that this is only the case when the prior beliefs themselves were arrived at by a normative process—which they consider unlikely to be the case.
readiness of their subjects to use a body of evidence--already processed with this "justifiable bias"--to strengthen the very prior belief that led to the processing bias to begin with is problematic. The nature of this bias in evidence evaluation would necessarily result in setting up a mechanism in which the confidence with which one holds a prior belief could not be lessened--let alone falsified--unless presented with undisputable contradictory data. Thus while some prior belief influence on the evaluation of evidence may be justified, it seems clear that the extent to which this influence is exerted can lead to situations where prior belief becomes remarkably undisturbed by consideration of all the evidence.

The phenomenon of belief perseverance (e.g., Anderson, 1983; Anderson, Lepper & Ross, 1980) makes a particularly strong case for the the nonnormative influence that prior belief can have on reasoning. Using an experimental paradigm that induces a belief, Anderson and his colleagues have repeatedly shown in a series of studies that participants' reasoning continue to be influenced by the induced belief even after the original evidence used to induce the belief was thoroughly discredited. It would seem in this case difficult to justify this prior belief influence as it is not the result of an appreciable base of relevant knowledge on some subject, but rather, a belief resulting from a small amount of information. That the induced belief persists to be manifest in reasoning behaviour after discrediting the initial information having led to that belief to begin with should at the very least suggest the possible nonnormative nature of belief bias.

**Addressing the Status of Normative Models**

A strategy outlined by Stanovich and West (1998a) to address the normative issue will be adopted for the present research. The strategy takes as a starting point the notion that human cognition is remarkably adaptive as evidenced in its enormous success. This view is is particularly salient in the work of Anderson
Belief Bias

(1990) whose proposed General Principle of Rationality states “The cognitive system operates at all times to optimize the adaptation of the behavior of the organism” (p.28). A real concern then is that any deviations of the modal response from the dictates of the normative models may speak more to the inadequacy of the normative models rather than human rationality itself (see Broome, 1990; Einhorn & Hogarth, 1981; Gigerenzer, 1991, 1993, Stanovich & West, 1998a; Stanovich, in press; Stich, 1990). These critiques offer a different interpretation to the classic research literatures on heuristics and biases (e.g., Kahneman & Tversky, 1972, 1973, 1979, 1982, 1984) and social biases (e.g., Nisbett & Ross, 1980; Nisbett, 1981). Although some advocates of the appropriateness of normative models as optimal standards of rationality do not feel the need for them to be constrained by descriptive accounts (e.g., Baron, 1985a, 1988), the sort of criticism spelled out above holds that descriptive data is relevant to the choice of normative models.

Stanovich and West’s (1998a) insight was to address the issue by using an informative analysis of individual differences. Specifically, there is large variation found in the descriptive data itself, and importantly, some individuals within this descriptive data do give the normative response on all the rational reasoning tasks. The strategy is then to determine if these individuals who do supply the normative response are simply a natural consequence of statistical “noise” around the modal (and truly optimal) response in the descriptive data or whether these individuals are distinguishable by factors that may have some bearing on how we view their selection of the normative response. What Stanovich and West (1998a) found is that the subgroup of participants who do give the normative response are indeed differentiated from those individuals giving the modal response on a factor that we might want to take into consideration—they are more intelligent. Individuals giving the normative response on several
tasks of decontextualized reasoning used by Stanovich and West (1998a) score significantly higher on traditional measures of psychometric intelligence than those participants who gave the modal response.

Under the assumption that descriptive data does have some bearing on our evaluation of the status of normative models, the question that arises then is the following: Since there is large variation in this descriptive data, which portion of the descriptive data is most relevant to the selection of normative models? Is our selection of normative models to be particularly influenced by the modal response, or the response given by those found to have the highest levels of psychometric intelligence? Since intelligence is almost universally viewed as a significant determinant of cognitive competence (Anderson, 1990, 1991; Baron, 1985a, 1985b; Carroll, 1993, 1997; Gordon, 1997; Gottfredson, 1997; Hernstein & Murray, 1994; Hunt, 1995a, 1995b; Larrick, Nisbett & Morgan, 1993; Oaksford & Chater, 1993, 1994, 1995; Shanks, 1995; Sternberg, 1985; just to name a few), it would be difficult to reconcile this view with the fact that those with the highest levels of cognitive capacity are less likely to be effective reasoners than individuals possessing lower levels of cognitive capacity. To summarize the critical point made by Stanovich and West (1998a), insofar as we use descriptive data to constrain normative models, it would seem that consideration should be given to all aspects of the descriptive data, including covariance patterns in addition to central tendencies.

**Ecological Validity and the Projection of Belief**

Assessing an appropriate level of prior belief influence on the evaluation of data is problematic and some important insights can be discerned from consideration of judgments or reasoning executed in information environments which are ecologically valid. For many of the judgments we make concerning evidence quality, our prior beliefs are to some degree diagnostic. Kornblith (1993) builds upon this point to argue “...If, overall, our belief-generating mechanisms
Belief Bias

21
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). Thus one concern may be that demonstrations of human irrationality result from experimental conditions artificially invalidating informational cues which are both valid and useful in the natural ecology but detrimental to the contrived experimental reasoning task at hand (see Gigerenzer, 1993).

A potentially important lesson can be drawn here from the social perception literature on the "false consensus effect" observed in opinion prediction tasks (Marks & Miller, 1987; Mullen, Atkins, Champion, Edwards, Hardy, Story, & Vanderklok 1985). The false consensus effect refers to the reported tendency of people to project their own personal opinions, attitudes, behaviours, etc. when predicting those of others. Thus, for example, participants who endorse the statement “Television is my primary form of entertainment” are likely to also project this opinion in their prediction of what others would think (viz., by projecting consensus). Similarly, those that do not endorse the above statement also tend to project their own opinion in their prediction of how others will answer. The initial interpretation of this phenomenon was to see it as a nonnormative tendency in reasoning—a sort of egocentric bias. Further to this, it would appear to be a bias immune to critiques of failing to use ecologically valid stimuli since the elicited predictions concern the participant’s surrounding social environment. It has been persuasively argued, however, that this observed response tendency may not be nonnormative after all.

Using a Bayesian analysis, Dawes (1989, 1990) demonstrated that belief projection can be normally justified. This is the case because by mathematical definition of majority, a given individual is in the majority most of the time. Thus an individual’s own opinion about some issue is usually a diagnostic cue for the general consensus of the population. Similarly, employing Brunswikian
analyses, Hoch (1987) argued that one crucial factor to consider in opinion predictions tasks is the consequence of belief projection on predictive accuracy—something not addressed in the traditional between-subjects analyses. Specifically, are those participants who do not project belief any more or less accurate than those participants who do project belief? The traditional thinking on this was that belief projection was nonnormative. Hoch (1987), however, obtained results suggesting that projecting belief improves predictive accuracy. That belief projection in a ecologically valid task actually facilitates predictive accuracy undermines the entire notion that the false consensus effect is a nonnormative response tendency.

As an exploratory investigation, the present research will also investigate the normative appropriateness of belief projection, and more importantly, the correlates of belief projection. This will be done by administering the classical opinion prediction task paradigm as well as another analogous task designed for this research. The inclusion of social perception tasks is driven by the rationale that prior belief influences on reasoning and judgment may also be viewed as “projecting” beliefs onto the world. Thus the correlations between performance on the prior belief tasks used in this research and performance on the social perception tasks is of interest.

The Domain Specificity and Generality of Belief Bias

The ability to avoid undue influence of prior beliefs on reasoning has been repeatedly stressed as a desired trait of a critical thinker by several researchers. Thus for example, Nickerson (1987) argues that the critical thinker appreciates “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). Norris and Ennis (1989) argue that an important aspect of critical thinking is the ability to “reason from starting points with which they disagree
without letting the disagreement interfere with reasoning" (p. 12). Similarly Paul (1990) states that what he calls “strong-sense” critical thinkers are able to “enter sympathetically into, and reconstruct, the strongest arguments and reasons for points of view fundamentally opposed to their own” (p. 6) and goes on to state that they are less likely be “routinely blinded by [their] own point of view” (p. 7). Several other researchers also argue that this ability to avoid belief bias type of influences is an important general trait of critical thinking (e.g., Baron, 1991, 1995; Klaczynski, 1997; Klaczynski, Fauth & Gordon, 1997; Kuhn, 1991, 1996; Perkins, 1995; Perkins, Jay, & Tishman, 1993, Stanovich & West, 1997, 1998a).

A crucial aspect of this research tradition in critical thinking is an underlying assumption that critical thinking skills cut across different content domains. Discussions in the critical thinking literature, whether explicitly or implicitly, tend not to constrain critical thinking ability to specific content domains. Characteristics of critical thinking such as reflectiveness, fairness, and objectivity, for example, do not have their scopes restricted to a specified domain, but rather styles or dispositions promoting these characteristics are outlined as domain general. Baron (1985b) makes the domain generality assumption explicit: “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). Despite this underlying assumption in much of the critical thinking literature, there has been virtually no multivariate investigations to show that this indeed is the case. In addition, other research traditions have not adopted this domain general view of critical thinking.

The last decade has seen the preponderance of the relevant developmental research emphasizing the importance of domain specific knowledge as the determinant of good thinking. Perhaps the most notable example of this is found within the position articulated by situated cognition theorists (e.g., Brown,
Collins and Duguid, 1989) in which cognition is viewed as heavily reliant on the specific context in which it occurs. Also working within the developmental field, Ceci (1993, 1996) has focussed attention on demonstrations in which formally identical tasks elicit varying performance levels across differing contexts. This premium placed on context is strongly expressed in the work of Stigler and Baranes (1988): “What differentiates subjects who exhibit expert performance from those who perform at a lower level is not the presence of higher levels of general cognitive abilities, but rather is a greater accumulation of knowledge specific to the domain of expertise” (p. 257). Thus these contextualist positions of cognition do not accept the view that critical thinking skills are domain general--context-free thought simply does not exist. The contextualist would not predict that belief bias is an individual difference variable cutting across various domains. Instead an extreme contextualist position would predict that there will be no cross-domain correlation between belief bias indices.

**Addressing the Domain Specificity and Generality Issue**

Belief bias has been extensively studied in the domain of syllogistic reasoning (see Evans, Over & Manktelow, 1993), and although it has been demonstrated in informal reasoning tasks as well (e.g., Baron, 1995; Kuhn, 1991), investigations of belief bias have largely been undertaken within the domain of verbal reasoning. One aim of the present research was to examine whether belief bias can be described as an individual difference variable cutting across distinctly different judgment domains or whether it is specific to particular content domains. Specifically, investigations of the two general types of tasks that have been the focus of this literature (deductive reasoning and informal reasoning) have repeatedly demonstrated large individual differences in the degree to which people's judgments are influenced by their prior belief. Yet few previous studies have made an attempt to examine the covariates of these individual differences,
nor have previous studies examined whether they were reflecting a domain general tendency to project prior belief onto evidence. To address this issue, a multivariate investigation was conducted in order to assess the extent to which belief bias covaries in distinct content domains of reasoning and judgment. The content domains investigated here were as diverse as deductive reasoning and perceptual judgment. The perceptual judgment task used here extends the domain of investigations of belief bias by examining reasoning in an area quite different from the verbal reasoning situations which have been the focus of investigation. In addition to the fact that the processing demands of this perceptual task are radically different from the typical reasoning tasks used to demonstrate belief bias, another very nice aspect of the perceptual task used in this research is that it is structurally analogous to the syllogisms discussed above. Furthermore, like the syllogistic arguments, this perceptual task affords a direct measure of belief bias.

The perceptual judgment task used in this study was a height estimation task adapted from the work of Nelson, Biernat, and Manis (1990). Participants in the Nelson et al. (1990) study were required to judge the height of photographed people. The results showed that height judgments were related to the actual target heights, but that these judgments were also systematically biased by a prior belief—viz., a gender/height relationship. Male participants were judged to be taller and female participants shorter than was really the case. Thus, although gender is a diagnostic cue for height judgments (on average males are taller than females), the tendency was to over-project this cue. Of particular interest to the present research is that this over-projection was maintained in a condition where the male and female pictures were matched in height and the participants were told of this and explicitly warned not to be influenced by the gender of the target.
The logic behind this latter condition closely mirrors the logic of the reasoning tasks demonstrating belief bias in syllogistic reasoning. In both cases, the participant is told to ignore a prior belief that normally may be diagnostic but is currently rendered nondiagnostic for the task at hand. As true with demonstrations of belief bias in syllogistic reasoning, participants in the Nelson et al. investigation—at least in the aggregate—are unable to ignore the nondiagnostic prior belief (viz., the gender/height relationship in this case) and continue to project the gender cue in their judgments. Nelson et al. (1990) did not investigate individual differences in the extent to which participants projected their prior belief. Analyses of individual differences in this task will be of central importance to this study.

Two versions of the Nelson et al. (1990) height judgments paradigm are incorporated into this study. In the first version—termed ecological height judgment task—the diagnosticity of the gender/height relationship was undisturbed by having participants view pictures of people who were a reasonably representative sample of the actual heights of males and females in the North American population. In the second version—termed the matched height judgment condition—participants viewed a sample of pictures where the mean heights of the males and females were matched and participants were explicitly made aware of this matching process (i.e., participants were informed that the gender cue was rendered nondiagnostic). This matched condition is structurally analogous to the syllogistic reasoning situation and also affords a direct index of belief bias.

Whereas the matched condition provides a perceptual task yielding both a structure and belief bias index that is directly comparable with a deductive logic task, viz., syllogistic reasoning, the strength of the ecological condition lies with its use of ecologically valid stimuli. The results of the Nelson et al. (1990) study
demonstrated that participants over projected the gender cue in this ecologically
valid task. Participants thus appeared to be overly influenced by their prior belief.
What was not investigated, however, was the consequences of projecting the
gender cue, viz., the prior belief. It is at this juncture where investigations of the
false consensus task become particularly relevant. The so-called false consensus
effect has been shown to be somewhat of a misnomer since projection of belief
actually is found to facilitate predictive accuracy. An identical analysis also
needs to be performed on participants' “overuse” of the gender cue. Taking a
lesson out of the false consensus literature, an apparent overuse of a diagnostic
cue (viz., own opinion) still resulted in facilitating predictive accuracy. The
present study will also explore the performance on the ecologically valid height
judgment task in the context of the these issues raised in the social perception
literature.

Predicting Belief Bias Using a Cognitive Capacity -
Thinking Dispositions Framework

Given that there are individual differences in the extent to which people
exhibit belief bias, are there any individual difference measures that would allow
us to predict (with some degree of confidence) the magnitude of belief bias for any
given individual? Addressing this question is nothing less than constructing a
partially explicit model of some mechanisms subserving belief bias. An obvious
place to first start searching for factors that have some bearing on the observed
individual differences in belief bias is the one individual difference variable that
has received the most research attention in the entire history of psychological
inquiry: Intelligence.

As already outlined above, adaptionists would certainly want to argue that
individuals possessing higher levels of Intelligence are more likely to be “effective
across environments” (Anderson, 1990, 1991; Larrick, Nisbett & Morgan, 1993;
underlying rationale here is that individuals with higher levels of intelligence are more likely to employ the normative reasoning strategies. This increased likelihood of utilizing normative strategies can be due to at least two reasons. First, normative strategies may sufficiently tax cognitive resources to a degree where only those individuals with higher levels of cognitive capacity may be capable of employing them. Second, even in the case where normative strategies do not heavily tax cognitive resources, their effectiveness (and consequent employment) may only be recognised by those individuals with higher levels of cognitive capacity (see Stanovich & West, 1998a).

Scores on traditional measures of psychometric intelligence have been repeatedly found to predict performance in school and the work place--their success in predicting cognitive competence is undeniable. (Carroll, 1997; Gottfredson, 1997; Lubinski & Humphreys, 1997). Cognitive capacity is operationalized in this study in terms of performance on traditional tests of psychometric tests such as Ravens Progressive Matrices and subtests from the WAIS-R. Performance on these tests load heavily on psychometric g (Carpenter, Just, & Shell, 1990; Carroll, 1993, 1997; Matarazzo, 1972) and have also been linked to both neurophysiological measures and information processing indicators of efficient cognitive computation (Caryl, 1994; Deary, 1995; Deary & Stough, 1996; Detterman, 1994; Fry & Hale, 1996; Hunt, 1987; Stankov & Dunn, 1993; Vernon, 1991, 1993).

Although cognitive capacity has some real world bearing on cognitive competence, it nevertheless leaves much of the variance unaccounted. This is particularly evident in empirical investigations using various rational thinking tasks (Baron, 1985a, 1988; Klaczynski, Fauth & Gordon, 1997; Perkins, Farady, and Bushey, 1991; Stanovich, 1994; in press; Stanovich & West, 1997, 1998a). That intelligence, or psychometric intelligence to be more precise, leaves much
unexplained in matters of thinking and reasoning has, of course, not gone unnoticed. Several proposals of theory and research endeavors can be understood as attempts at developing more powerful frameworks in accounting for "good thinking" than that afforded by considerations of psychometric intelligence alone. This is especially true within the interdisciplinary field of cognitive science. Whether as a consequence, an impetus or a pure epiphenomenon of cognitive science, the conception of intelligence migrated from a heavy "intelligence as structure" view to an intelligence as "process" view. Specifically, several formulations within cognitive science theory have sought to explicate factors leading to the process of good thinking or what we might call here critical thinking. For present purposes it will suffice to spell out a broad characterization of good thinking as what Jonathan Baron (1988, p.28) defines as "whatever helps us fulfill our personal goals". Even the word "intelligence" in cognitive science is often replaced by new terminology having fewer of the more dated connotations such as cognitive ability or cognitive capacity. Except for a few instances were historical usages are more convenient, the terms cognitive capacity and cognitive ability will be employed henceforth to denote that which is also implied by psychometric intelligence.

Expanding upon Baron (1985a, 1988), Stanovich and West (1997, 1998a; see also Stanovich, in press) effectively argue with data that a significant portion of the variance unaccounted for by cognitive capacities may be accounted for by thinking dispositions. Various terminology has been employed surrounding proposals similar to the idea of thinking dispositions. Thus, for example, while Baron (1985a, 1988) speaks of thinking dispositions (as also does Ennis, 1987), others employ terms such as intellectual styles (Sternberg, 1988, 1989), cognitive emotions (Scheffler, 1991), habits of mind (Keating, 1990, 1996; Margolis, 1987, 1993), inferential propensities (Kitcher, 1993, pp. 65-72), epistemic motivations
Belief Bias

(Kruglanski, 1990), constructive metareasoning (Moshman, 1994), and cognitive styles (Messick, 1984, 1994). All these formulations can be viewed as partially motivated by varying degrees of dissatisfaction with the singular use of cognitive capacity as the individual difference variable accounting for the observed variance in rational thinking.

Drawing on Baron’s (1985a, 1988) work, Stanovich and West (1997, 1998a) outline an important distinction between cognitive capacities and thinking dispositions. Thinking dispositions relate to the adequacy of belief formation and decision making. Examples include the disposition to search selectively for evidence supporting a present belief, the disposition to place much weight on evidence consistent with a prior belief or little weight on inconsistent evidence. These particular dispositions may serve to maintain a favoured belief at the expense of a fair and impartial evaluation of the available evidence. Certain thinking dispositions may also promote a more data sensitive and normatively appropriate analysis, such as dispositions which result in more open-minded approaches or a disposition not to easily give up on a problem when difficulty is encountered (Baron, 1985a; 1988; Stanovich & West, 1997, 1998a).

As Stanovich and West (1997) argue, the proposed distinction between cognitive capacity and thinking dispositions can be grounded in an existing framework developed by computational theorists. Cognitive capacity and thinking dispositions may be seen as exerting influence at different levels of a cognitive theory. Distinctions between these explanatory levels have been previously articulated in the levels of analysis approach to cognitive theory (see Marr, 1983; Newell, 1982, Pylyshyn, 1984, Anderson, 1990). Although the choice of terminology used by these theorists are distinct, they share much in common. My discussion will focus on Anderson (1990) who describes four separate levels of analysis in cognitive theory.
Starting from the bottom and working up, the first level described by Anderson (1990) is the biological level. Although all of cognition is somehow ultimately anchored upon the activity of its physical implementation at this bottom level of analysis, there simply is not enough known about the brain to specify this biological level. Thus the next level up in cognitive analysis, which Anderson terms the implementation level, serves as a useful computational approximation for the biological level and a holding position until presumably a mature neuroscience is in a position to specify the actual biological level. The next two levels of analysis are particularly relevant to the distinction Stanovich and West (1997) make between cognitive capacity and thinking dispositions. Anderson terms the next level of analysis as the algorithmic level. This level specifies how a cognitive task is accomplished by specifying the type of procedures or steps (viz., algorithms) that must be executed. This is where the construct of cognitive capacity is proposed to have its influence on reasoning performance. Cognitive capacity will set limitations upon the complexity of procedures or steps that can be supported by the cognitive apparatus. Working memory capacity, perceptual speed, discrimination accuracy, and efficiency of information retrieval from long-term memory are all quantifiable dimensions of relevance to information processing theory, and importantly, in which individuals are found to vary substantially. These are furthermore the underlying factors of cognitive capacity setting the upper boundaries on processing efficiency at the algorithmic level of analysis.

Anderson's final and top level of analysis for cognitive theory, the rational level, can be viewed as the province of thinking dispositions. The rational level specifies the goals of the computation. At this level of analysis the system can be understood in terms of what it is attempting to achieve and why. In as far as thinking dispositions (as will be outlined here) can be taken as an individual's
indication of epistemic goals relevant to thinking and reasoning, measures of thinking dispositions can be interpreted as a proxy for some of the goals and epistemic values at the rational level of analysis.

In order to illustrate this connection between the rational level of analysis and thinking dispositions, it would be useful to outline a dispositional scale developed and used in this research and demonstrate why it can be taken to serve as a proxy for some important goal or epistemic value dimension at the rational level of analysis. The belief identification scale was inspired by a theoretical paper from Cederblom (1989), in which he argues that a potential dimension of importance in explaining how willing people are to reason about issues involves how closely people identify their beliefs with their concept of "self”. At one extreme of this dimension, Cederblom argues, are people who closely identify their beliefs with their own sense of persona. These people may in consequence be less inclined to revise or change beliefs since they view them as such an intimate part of who they are. At the other end of the spectrum are those who see themselves simply as a belief forming-process. These people don’t identify their beliefs with their sense of self and simply view belief as external entities that may be evaluated, revised or discarded without impacting on their own persona. As a consequence of this favorable view, Cederblom argues, people would be less reluctant to depart with previously held beliefs as opposed to valuing the maintenance of existing beliefs. Kuhn (1992) has also emphasized the importance of this sort of dimension on peoples’ ability to engage in constructive argumentation. Referring to this dimension under the rubric of “belief ownership”, Kuhn makes the case that having a sense of “ownership” in your beliefs and thus creating a sort of vested interest in their continued existence can be detrimental to a objective or fair evaluation of their actual validity.

A nine item scale was distilled from the ideas expressed in Cederblom (1989)
to assess the possible relevance of this epistemic goal dimension on prior belief reasoning. An example of an item is "Someone who attacks my beliefs is not insulting me personally"--reverse scored. A person who scores high on this scale is indicating that the stability of their belief system is an important epistemic goal to them. With the attainment of a goal, further processing is less likely to occur. Goals held at the rational level would thus seem to have real implications for things such as when the processing of information is terminated or "stop rules" are invoked by the system. A system that has attained a valued goal, e.g., justifying a held belief in the presence of belief-enhancing evidence, may prematurely stop processing. The same system may continue to process in attempt to justify a belief in the presence of belief-threatening evidence.

Baron states that thinking dispositions are "learned tendencies to behave in certain ways" (Baron, 1991, p. 118). As such, Baron views thinking dispositions as particularly malleable--which would then of course have great relevance to education. Under the current formulation of thinking dispositions, however, the issue of malleability it is not quite clear since they are taken to be as largely epistemic goals held by a cognitive system. Common to both formulations is the crucial point that thinking dispositions determine how the cognitive capacities are put to use. Illustrating this point with a thought experiment, Baron (1985a) asks the reader to imagine that cognitive capacity (e.g., working memory capacity, processing speed etc.) could be increased through the administration of a harmless drug. Baron then asks the reader to consider what impact this increase in cognitive capacity would have on a reasoner. More specifically, would the administration of this drug increase the quality of reasoning? Baron argues that by-and-large it would not and suggests a biased processor of information (or closed-minded individual) would continue to be so upon administration of the drug--albeit a more efficient and faster biased processor of information. Perkins
et al. (1991) outline a similar argument in their distinction between cognitive capacities and thinking dispositions. Extending on the notion of traditional IQ tests tapping the “computational power” of the brain (see Cherniak, 1986), Perkins et al. suggest that “having a high IQ gives no guarantee of one’s using it well anymore than having a lot of horsepower under the hood of your car guarantees your driving the car well” (p. 103)

The Research Strategy

The general analytical strategy was to examine the covariance structure of direct indices of belief bias across tasks as diverse as deductive logic reasoning and perceptual judgments as well as indices of performance on three other tasks reflecting the ability to evaluate evidence and argument independent of prior belief. A perceptual judgment task pitting prior belief against accurate performance extends the domain of investigations of belief bias by examining it in an area quite different from the verbal reasoning situations which have been the focus of investigation. Additionally, the perceptual reasoning task contained two conditions—one in which prior belief contradicted the relationships in the stimulus set (a situation completely analogous to the syllogistic reasoning problems discussed above) and one in which prior belief was consistent with the relationships in the stimulus set, viz., an ecologically valid task. Also included in this research are two tasks in the tradition of the social perception literature. These tasks are designed to assess the extent of belief projection, and importantly, the consequences this projection has when done in ecologically valid environments. Also to be explored in this covariance structure are measures of general cognitive ability (using some traditional tests of psychometric intelligence) and thinking dispositions.

The thinking disposition scales used in this research were chosen because they embody the fundamental assumption of the critical thinking literature
reviewed above, i.e., that there are domain general dispositions toward critical thinking (Baron, 1993, 1994; Perkins et al., 1993). Thinking disposition scales incorporated into this study were selected on the basis of appearing to have epistemic significance (see Cederblom, 1989), for example: the disposition to search selectively for evidence supporting a present belief; the disposition to place much weight on evidence consistent with a held belief or little weight on inconsistent evidence. These dispositions relate to the adequacy of belief formation and decision making--they are measures of dispositions toward what Baron (1985a, 1988; see also Stanovich & West 1997, 1998a) call actively open-minded thinking.

Whereas the covariance structure between the indices on the prior belief tasks are directly relevant to the issue of domain specificity and generality of belief bias, the covariance of these indices with cognitive ability measures will address the normative issue. The critical thinking tradition holds that belief bias is a nonnormative tendency, as such belief bias should correlate negatively with measures of cognitive competence, i.e., cognitive ability measures. Furthermore, through the use of hierarchical regressions, residual performance variation on the prior belief tasks once cognitive ability is statistically controlled will be linked to actively open-minded thinking dispositions.

Hypotheses

Much of the present research is exploratory in nature. This is especially true with the nature and extent of the relationships existing between the various thinking disposition scales and task performance. Setting aside this large exploratory component to the present investigation, the rationale of the present research is very much grounded in the critical thinking literature. Specifically, belief bias is interpreted in the present study to be a nonnormative tendency. Under such an interpretation, the following hypotheses are made:
1. Belief bias indices will correlate negatively with cognitive ability on prior belief tasks in which prior belief is known by the subject to lack diagnosticity.

2. The covariation pattern among the prior belief task indices will demonstrate some domain generality in belief bias.

3. Belief bias and or the ability to reason independently of belief will correlate with thinking disposition measures.

4. Thinking disposition measures will account for additional variance on prior belief task performance after all the variance accounted for by cognitive ability is statistically controlled.

Some of the central concerns of critical thinking research will be addressed by the present research. Addressing the normative issue of belief bias is directly relevant to the conjecture that critical thinkers should avoid undue influence of prior belief on reasoning. That critical thinking skills cut across different content domains is directly tested in this research by comparing intercorrelations of performance on prior belief tasks drawn from domains as diverse as deductive logic reasoning and perceptual judgment. Furthermore an explorative endeavour to identify thinking dispositions lending themselves to critical thinking on prior belief reasoning is also undertaken by the present research.
Method

Participants

A total of 126 participants (54 males and 72 females) recruited through poster advertisements distributed on campus of a large southern Ontario university. Multivariate outliers identified by using the Mahalanobis distance metric diagnostic were removed if the obtained value was greater than 4.0. This resulted in the eliminating 2 participants from the study thus bringing the total number of participants to 124. All but 3 of these participants were undergraduate students at the university in which the study was conducted. Two of these three in question were graduate students. One was a history major attending the same university and one was a psychology major (history and systems specialist) attending another university in the same city. The third non undergraduate participant was a student enrolled in a technology program at a local community college. The average age of the participants was 21.9--this, however, was highly skewed due to a small group of mature students. The modal age was found to be 19. For the 122 participants who volunteered high school grade information on a demographics form, the mean reported final high-school year average was 84.3% (SD = 6.20). The current mean reported university average was 73.4% (SD = 6.9) for the 119 participants who volunteered this information. The study attracted participants from diverse academic backgrounds such as social science--48 participants (38.7%), science and/or math--36 participants (29%), the arts--21 participants (16.9%), business related programs--7 participants (5.6%), 3 participants (4.9%) were enrolled in other programs and the remaining 9 participants (7.3%) were undecided about a major. Participants were also somewhat fairly distributed in their program year: 37.1% of the participants were first year students, 21.0% were second year, 16.9 third year, 17.7 fourth year
while the remaining 7.3% reported to be in year 5 or over. Participants volunteered for a study which they were informed concerned human judgment and reasoning. Participants were also informed that the testing would take about 3 hours and 15 minutes for a total of $20 payment.

**Prior Belief Reasoning and Judgment Tasks**

**Syllogistic Reasoning Test.** 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 (for example, All living things need water; Roses need water; therefore, Roses are living things--which is invalid). Eight of the problems were worded such that the validity judgment was congruent with the believability of the conclusion (for example, All fish can swim; Tuna are fish; therefore, Tuna can swim--which is valid). Eight of the problems involved imaginary content (for example, All opprobines run on electricity; Jamtops run on electricity; therefore, Jamtops are opprobines--which is invalid).

Participants were instructed 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 subjects indicated their responses by circling one of the two alternatives: a. Follows Logically,  b. Does Not Follow Logically. Some examples of the syllogisms used in this study can be found in Appendix A.
There are several ways to derive a response measure with this task. Of particular importance is a direct index of belief bias which can be obtained by subtracting the score of the incongruent syllogisms (minimum score of 0 and maximum of 8) from the score of the congruent syllogisms (minimum score of 0 maximum of 8). The resulting belief bias index score may range from -8 to +8. Positive scores indicate belief bias with the higher scores indicating greater levels of belief bias.

Height Judgment Tasks. This study employed the height judgment task paradigm used by Nelson et al. (1990). Participants were presented with full length pictures of seated males and females. The participants task was to judge as accurately as possible the height of each of these pictured models. Participants were encouraged to be as accurate as possible and not let the gender of the target influence their judgment. Two slightly different versions of this task were used. In the first version, participants were simply told to estimate the height of several individually presented models after reading the instructions that follow:

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 in feet and inches. 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
Participants then went on to estimate the height of 83 pictured models (40 male and 43 female). Some examples of the stimuli used in the height judgment task can be found in Appendix B. Pictured models were selected in such a manner as to be ecologically valid in terms of height. Specifically, the average male and female height was set at 5'10 1/2” and 5'5 1/2” respectively which is in accordance with average heights for both sexes (Dairy Council Digest, March-April 1980). This version will be referred to as the Ecological Height Judgment Task henceforth.

The administration of the second version of the height judgment task directly followed the first version. This version forewarned participants that in the stimulus materials to be judged the generally accurate stereotype that males are taller than females did not hold. This important point was made via the instructions that follow:

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.**

In effect, Participants were informed that the gender cue was rendered nondiagnostic and consequently should be ignored in their ensuing height
judgments. Thus the logical structure of this task is completely analogous to the syllogistic reasoning test—in both tasks participants are explicitly instructed not to project their nondiagnostic prior beliefs in their reasoning and judgment activities. Participants estimated the heights of 46 pictured models (23 male and 23 female). The average male and female matched heights was set at 5'8 1/2". This version will be referred to as the Matched Height Judgment Task henceforth.

Once again there are several ways to derive a response measure for both the Ecological and Matched height judgment tasks. One effective index that will be used for both versions is simply to compute for each participant the point biserial correlation between the target gender (G; scored a 0/1) in the photographs and the participant's height estimation (E). A useful notation for this index is \( r(G,E) \). Ideally the correlation between the target gender cue and a participant's height estimations, i.e., \( r(G,E) \), in the Matched condition should be zero.

While the \( r(G,E) \) index in the Matched condition serves as a useful index for the degree to which a participant projects the nondiagnostic gender cue (i.e., a prior belief), it does not take into account the differential difficulty imposed on participants by varying degrees of prior belief in the magnitude of this gender-height difference. Participants who believe, for example, that males are 6 inches taller than females have to adjust their responses to a greater degree than those participants who believe the difference to be only 1 inch. A measure of the participants prior belief about the magnitude of the difference between the gender-height difference can be obtained from the the ecological set: The mean male height estimations minus the mean female height estimations. i.e., \((M-F)_{\text{ecological}}\). This prior belief estimate was used to construct an index of the degree of adjustment from the ecological set that was achieved in the matched set. This index, \( HJ \text{ Adjustment} \), was simply \((M-F)_{\text{ecological}}\) minus \((M-F)_{\text{matched}}\).
viz., the mean difference between male and female height estimations in the ecological set minus the mean difference between male and female height estimations in the matched set. For purposes of clarity and efficiency, a discussion of other relevant obtained indices will be outlined in the results section.

**Argument Evaluation Test.** The argument evaluation test (AET) is an instrument developed by Stanovich and West (1997) to assess prior belief influences on reasoning in the verbal argument domain. The AET consisted of two parts. In the first, participants indicated their degree of agreement with a series of 23 target propositions on a four-point scale: strongly agree (scored as 4), agree (3), disagree (2), strongly disagree (1). 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, in one item, the target proposition was: The welfare system should be drastically cut back in size. Scores on this part of the questionnaire will sometimes be 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 1.85 for the item “It is more dangerous to travel by air than by car” to a high of 3.66 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 subjects 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 subjects 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 provides a justification for the belief (in this case, for example, "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 subject is told to assume that the facts in the counter argument are correct. Finally, Dale attempts to rebut the counter argument (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 subject is told to evaluate the strength of Dale's rebuttal to the critic's argument. The instructions remind the subject that he/she is to focus on the quality of Dale's rebuttal and to ignore whether or not he/she agreed with Dale's original belief. The subject then evaluates the strength of the rebuttal on a four point scale: very strong (scored as 4), strong (3), weak (2), very weak (1). 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 1.85 to 3.68 across the 23 items.

The analysis of performance on the AET required that the subjects' evaluations of argument quality be compared to some objective standard. The authors of the AET employed 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 AET authors judged the strength of the rebuttals. The median correlation between the judgments of the eight experts was .74. Although the problems were devised by the two authors, the median correlations between the authors' 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 given the item by the subjects was 2.00 (weak) although the participants’ mean prior belief score indicated a neutral opinion 2.52. As was the case in past investigations (Stanovich & West, 1997, 1998), efforts to construct items so that objectively strong and weak arguments were equally associated with the prior beliefs that participants would endorse continued to be successful, as there was a nonsignificant .14 correlation between the sample’s mean prior belief scores and objective argument quality.

Individual differences in participants’ relative reliance on argument quality and opinion agreement were examined by running analogous regression analyses for each participant’s individual responses. That is, a separate regression model was constructed for each participant by simultaneously regressing his/her evaluations of the rebuttals on the experts’ ratings of argument quality and the participant’s level of agreement with the target opinion, i.e., a multiple regression with a participant’s argument evaluations as the response variable and both the expert ratings and respective prior belief
scores as the predictor variables). Thus, for each participant, these analyses resulted in a separately estimated beta weight for the experts' opinions and for the participant's level of agreement with the item. The former beta weight—an indication of the degree of reliance on argument quality independent of the participant's own opinion—is the primary indicator of the ability to evaluate arguments independent of one's beliefs.

The AET version used here was Canadianized from the original American content version. Modifications involved the replacement of a given American term with its Canadian counterpart or some minor rewording. Some examples of these changes include “The national debt should be reduced by cutting the Congressional salaries” changed to “...reduced by cutting the salaries of Members of Parliament” and “Labor unions are a major cause of the decline in the U.S. economy” was changed to “...major cause of the decline in our economy”. All modifications did not change or alter the logical form of the original argument. AET items numbers modified from the original were 10, 11--which also required the prior belief score to be reverse scored, 14, 17, 19, and item 20. The instructions and some examples of items from the AET are presented in Appendix C.

Covariation Judgment Task. The covariation task employed here involved an adaptation of an existing paradigm where people are presented with covariation information that is accommodated by the format of a 2 x 2 contingency table (see Wasserman, Dorner, & Kao, 1990) with the addition of a belief-bias component to it (see Evans, Over, & Manktelow, 1993; Levin, Wasserman, & Kao, 1993). This task is modified from a belief bias version designed and used by Stanovich and West (1998a).

The current task has been structured analogously to the above AET. The covariation judgment task, like the AET, consists of two separate forms which
were separated by other tasks. Participants essentially evaluated 25 contingencies which were embedded within the context of 25 different hypothetical relationships. Each of the 25 problems had two parts. On the first form, participants were asked for their opinion on several sets of hypothetical relationships between two variables. Once again, scores on this part of the questionnaire will sometimes be referred to in an abbreviated fashion as prior belief scores. For example, in one problem the subjects were asked whether they believed that couples who live together before marriage tend to have successful marriages. They indicated their degree of agreement with this hypothesized relationship on a scale ranging from -10 (strongly disagree) to +10 (strongly agree) and centered on 0 (neutral). Participants indicated their opinion on a total of 25 sets of hypothetical relationships. The second form asked participants to evaluate the degree of association between the same 25 sets of two variables in the data of a hypothetical research study. For example, for the marriage problem mentioned above, participants were told to imagine that a researcher had sampled 250 couples and found that:

- 50 couples did live together and had successful marriages
- 50 couples did live together and were divorced
- 50 couples did not live together and had successful marriages
- 100 couples did not live together and were divorced

These data correspond to four cells of the 2 x 2 contingency table (see Figure 1) traditionally labeled A, B, C, and D (see Levin et al., 1993). Subsequent to the presentation of the 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).
Figure 1. Illustration of the four resulting cells of numerical data conforming to a 2x2 contingency table.

The remaining 24 problems dealt with a variety of hypotheses (e.g., that secondary smoke is associated with lung problems in children; that exercise is associated with a sense of well-being; that eating spicy foods is associated with stomach problems; that being an early-born child is associated with high achievement; that getting chilled is associated with catching a cold; that psychics can help police solve crimes; that watching violent television is associated with violent behavior). Some examples of the covariation problems used in this task can be found in Appendix D.

The cell combinations used in the 25 problems were based on those listed in Table 2 of Wasserman et al. (1990). The cell values for a given data set in that table were randomly multiplied by 5, 10 or 15. The results of each of the hypothetical experiments were listed as above except that they were rotated in a fashion analogous to a Latin square. For example the initial problem was laid out in the cell order D-C-B-A, the next problem laid out in cell order C-B-A-D, followed by a B-A-D-C cell ordered problem etc.). This was done because participants, in a previous study by Stanovich and West were found to be highly attuned to the actual correlations present in the data. This modification was thought to make the task a bit more complex by changing the constant cell presentation order of D-C-B-A (as used by Stanovich & West) and thereby
eliminating a structural organization shared by all the problems. Other differences between the current version and the one used by Stanovich and West include the previously mentioned multiplication of Wasserman's original cell values by 5, 10 or 15--the previous version multiplied all cell values by five. This was done once again to increase the complexity of the task by eliminating a common order of magnitude differences in the data found across all the problems. The splitting of the task into a prior belief form and an evaluation form was yet another difference--the previous version consisted of only one form where participants first indicated their belief on some relationship and then immediately evaluated the relevant data. This split was accomplish to make the task less transparent. Also one item from original (#16) was dropped in order to eliminate a slight correlation found between the $\Delta p$ values and prior belief scores.

Previous experiments have indicated that subjects weight the cell information in the order cell A > cell B > cell C > cell D--cell D reliably receiving the least weight and/or attention (see Kao & Wasserman, 1993; Levin et al., 1993; Wasserman et al., 1990). The tendency to ignore cell D is nonnormative, as indeed is any tendency to differentially weight the four cells. The normatively appropriate strategy (see Allan, 1980; Kao & Wasserman, 1993; Shanks, 1995) is to use the conditional probability rule--subtracting from the probability of the target hypothesis when the indicator is present the probability of the target hypothesis when the indicator is absent. Numerically, the rule amounts to calculating the $\Delta p$ statistic: $[(A/(A+B))] - [(C/(C+D))]$ (see Allan, 1980). For example, the $\Delta p$ value for the problem presented above is +.167, indicating a fairly weak positive association. The $\Delta p$ values used in the 25 problems ranged from -.600 (strong negative association) to .600 (strong positive association).

As was the case in the AET, individual differences in participants' relative approximation of the $\Delta p$ statistic and opinion agreement were examined by
running analogous regression analyses for each participant's individual responses. Once again, separate regression models were constructed for each participant by simultaneously regressing his/her covariation judgments on the obtained $\Delta p$ values and the participant's level of agreement with the hypothetical relationship. Thus, for each participant, these analyses resulted in a separately estimated beta weight for the $\Delta p$ statistic and for the participant's level of agreement with the hypothetical relationship. The former beta weight—an indication of the degree of approximating the $\Delta p$ statistic independent of the participant's own prior belief—is the primary indicator of the ability to evaluate covariate relationships independent of one's beliefs.

**Confounded Variable Task.** This task was a modification to an experimental paradigm which has been extensively used by Schaller (see Schaller, 1992, 1994; Schaller, Boyd, Yohannes, & O'Brien, 1995, for review of the original experimental paradigm). For the purposes of this research, the task was substantially modified so as to pit prior belief against optimal response.

The initial stage of this task involved an attempt to induce a belief in the participant. Using a strategy successfully employed by Anderson's work in belief perseverance (see Anderson et al. 1980, Anderson 1983) the task begins by trying to induce the belief (via two concrete examples) that risk taking firefighters are generally more successful at their jobs than risk avoiding firefighters. Participants read two brief case studies. One case study outlined the successful rescue of a young boy by a risk taking firefighter. The second case study outlined the failure of a risk avoiding firefighter to save the life of a young boy under identical circumstances. The directionality of the chosen relationship takes advantage of the apparent general public opinion (i.e., a slight tendency to view risk taking personalities as characteristic of good fire fighters; see Anderson et al., 1980). After participants have read the two brief case studies, they are asked
to state whether they believe risk takers or risk avoiders are more successful as fire fighters. They are furthermore asked to state one good reason for their belief. The belief inducement was largely successful with 77% of the participants indicating that they believed risk taking firefighters to be more successful than risk avoiding fire fighters while the remaining 23% indicated the reverse. Participants also indicated on a six point scale the magnitude of their opinion on this real world relationship. The scale was anchored with “I strongly agree that risk-taking firefighters will be better performers”--scored as 1 and “I strongly disagree that risk-taking firefighters will be better performers”--scored as 6.

Immediately after this “belief inducement” stage, the participant is asked to assess the existing relationship in the data presented below:

<table>
<thead>
<tr>
<th></th>
<th>Total overall</th>
<th>Residential fire stations</th>
<th>Industrial fire stations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Succ.</td>
<td>Fail</td>
<td>Succ.</td>
</tr>
<tr>
<td>Risk avoiding</td>
<td>10</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Risk takers</td>
<td>15</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

A simplistic analysis of the above data (i.e., the judgment being based on only the overall summary portion of the table) suggests that risk takers make better fire fighters (a position aligned with the induced belief). This simplistic analysis, however, does not take into account the clear presence of a confound. There are proportionately more risk takers serving in residential fire stations, and furthermore, residential fire stations seem to be easier than industrial fire stations (as can be ascertained by comparing the success rates for residential versus industrial fire stations). Conversely, there are proportionately more risk avoiders in the difficult industrial fire stations. Thus if the participant reasons in a manner analogous to a partial correlation or engages in an intuitive ANCOVA-like reasoning about the presented data, it should be apparent that the data actually suggests--contrary to the belief-congruent and simplistic analysis--
that risk avoiding firefighters tend to be more successful than risk taking firefighters.

Having a large majority of participants endorsing the same directionality in the prior belief will make the statistical analysis and interpretation less cumbersome. Hence the use of the belief inducement stage from previous literature as well as taking advantage of the slight natural tendency found in general public opinion. Regardless of whether a participant indicates a prior belief that risk takers or risk avoiders make for more successful firefighters, the participant received confounded data in which the simplistic analysis is belief-congruent whereas the product of a more complex ANCOVA-like reasoning strategy would be a conclusion conflicting with the prior belief. This was accomplished via two reciprocal versions of the data evaluation instrument. The appropriate data set (instrument version) was given to a participant in accordance with their stated prior belief only after the prior belief has been indicated on the first instrument. Specifically, those participants who maintained that risk avoiding firefighters are generally more successful than risk taking firefighters--despite the belief inducement procedure to the contrary--were given the same data below with the exception that the data row corresponding to risk takers in the above table was assigned to risk avoiders and vice versa. The confounded variable task along with both data set versions (for the second instrument) are displayed in Appendix E.

Participants indicated their evaluation of the data on a 9-point scale anchored with “Risk taking firefighters are markedly superior to risk avoiding firefighters” scored as 1 and “Risk avoiding firefighters are markedly superior to risk avoiding firefighters” scored as 9. Thus participants who indicated the prior belief that risk takers are more successful as fire fighters than risk avoiders and yet respond at the higher end of this response scale--an evaluation that conflicts
with their prior belief--are engaging in ANCOVA-like reasoning. Conversely, scores on the lower end of this scale indicate that these participants have opted for a relatively simplistic analysis resulting in a belief congruent evaluation. The responses on this scale for those participants who indicated that risk takers are more successful fire fighters were reflected so that the directionality of the scale response was uniform across the two versions. Thus for all participants high scale scores indicate a simplistic analysis resulting in a belief-congruent evaluation of the data whereas low scale scores indicate a more complex ANCOVA-like analysis resulting in a belief-incongruent evaluation of the data. This derived index will be referred to as the belief congruence score henceforth.

Social Perception Tasks

Two further tasks were also included in the test battery. Although these additional tasks are not prior belief tasks in the manner outlined above\(^5\), they are relevant to addressing the consequences of projecting belief in reasoning and judgments. The two tasks below share the same structural logic, and performance on them can be indexed using identical methods. For purposes of clarity and efficiency, a discussion of the indices derived in these tasks will be outlined in the results section.

Opinion Prediction. Participants were presented with thirty statements used in previous consensus judgment research (e.g., I think I would like the work of a school teacher; My daily life is full of things that keep me interested; see Table 21). Sixteen items were taken from Dawes (1990), six from Hoch (1987), four from Sanders and Mullen (1983), and four from Campbell (1986). The task consisted of two phases separated in time by other tasks. In the first phase participants were

---

\(^5\) The following tasks differ from the prior belief tasks in two important features. First, there is no mention that a prior belief should be avoided. Second, no procedure was employed to invalidate the diagnosticity of prior beliefs, i.e., these tasks were performed within their ecologically valid environments. While some of the prior belief tasks also lacked one of these features (at least explicitly), no prior belief task lacked both of these features.
read the following instructions by the experimenter:

I’m going to read you several statements. After I read each statement I’ll ask you to estimate what percentage of the students participating in this study (all undergraduate U of T students like yourself) would agree with the statement (on a scale ranging from 0 percent to 100 percent).

Each statement was then read by the experimenter. The first statement was followed by the clarifying question “What percentage of the students participating in this study do you think agree with this statement?” Participants’ estimates were promptly recorded by the experimenter.

Later in the second phase of the task, participants were given a form which consisted of the same 30 statements. Participants read the following instructions:

Please read each of the following statements. Indicate your personal level of agreement (or disagreement) for each on the below scale.

For each statement participants indicated their degree of agreement on a four-point scale: strongly agree (scored as 4), agree (3), disagree (2), strongly disagree (1). The items elicited a wide range of levels of agreement and perceived consensus (see Table 21).

Knowledge Prediction. The Knowledge Prediction task was designed to be analogous in structure to the above Opinion Prediction task. In the first phase of the task, participants were asked to answer several general knowledge questions read out loud by the experimenter. The task began with the experimenter reading out loud the instructions that follow:

I’m going to ask you to answer some factual knowledge questions. They range from fairly easy to difficult. All questions require only one word answers. Guessing is encouraged when you are unsure about an answer. There is no time limit, but answer as quickly as
possible. If you don’t know, just say “pass” and I’ll move to the next question. Are you ready?

The experimenter scored the participants answer to each question as correct or incorrect. No feedback was given to the participant at this question-answer stage of the task.

In a second phase of the experiment, which immediately followed the first, participants were read the instructions that follow:

I’m going to read you all the questions again. This time I’ll ask you to estimate, for each question, what percentage (0 percent to 100 percent) of the students participating in this study, all undergraduate U of T students like yourself, you think will get each of the questions correct. For each question I’ll also let you know if you answered it correctly or incorrectly.

Participants were informed of the correct answers only after they gave their estimate. The only information which participants were provided with prior to each estimate was whether they answered the question correctly or incorrectly.

A total of 30 questions were drawn from a pool of normative data collected previously on 300 general-information questions (see Nelson & Narens, 1980). These questions covered a variety of topics associated with various levels of difficulty (see Table 22). The obtained probability of correct response associated with each item as reported in Nelson and Narens (1980) was utilized here as an index of item difficulty in the selection of the 30 questions used in this task. Easy, medium and difficult questions were equally represented in order to minimize any floor and ceiling effects in individual performance. For the same reason it was also determined that the questions should cover a wide variety of categories (viz., domains of interest). Not all the question categories in the original Nelson and Narens normative data were sufficiently distributed in difficulty level to
realize these two conditions. Thus the questions were drawn from five categories which did permit the above conditions: History, Geography, General, Nature and Entertainment. Six questions for each above category were drawn for a total of 30 questions. Within each of the 5 categories, 2 of the 6 questions were stratified into a further grouping level of Easy (probability of correctly answering around 0.75), another 2 into Medium (probability = 0.50) and the last 2 into Difficult (probability = 0.25).

Consideration was given to leave out any questions in which performance might be expected to significantly differ between Americans (the population in which this normative data was collected) and Canadians. For example, what is the capital of Canada (p = 0.104), who designed the American flag (p = 0.678). Also left out were potential chronologically sensitive questions such as who is Batman's Butler? The rationale here was that familiarity with Batman has reached a new high in recent years due to the new big-budget movies. Also avoided were questions where there was a comparatively large difference between males and females or between the two different universities in which the data was collected.

A further index was derived from this task to assess the normative issue. One point was given to each question answered correctly in this task to form an index of general knowledge. Thus the scores on this General Knowledge Test can range from a minimum of zero, i.e., no questions answered correctly to a high of 30, i.e., all thirty questions answered correctly.

**General Cognitive Ability Measures**

Participants completed a popular short form of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). The short form used consisted of the Vocabulary plus Block Design sub-tests. Previous research has shown that these two subtests attain moderate (Block Design) and high
(Vocabulary) correlations with the Full Scale, achieve consistently high reliabilities, and are good measures of $g$ (Sattler, 1988). Scores on these two subtests can be pro-rated to obtain a popular short form IQ score with high reliability and validity (Sattler, 1988). Using Sattler’s (1988) formulas a prorated IQ score for each subject was derived from the Block Design and Vocabulary subtests. This pro-rated IQ score will be termed general cognitive ability 1 (GCA1).

A second general cognitive ability measure (GCA2) was also derived by combining performance on a verbal and nonverbal test. The verbal measure was a brief vocabulary measure employing the checklist-with-foils format that has been shown to be a reliable and valid way of assessing individual differences in vocabulary knowledge (Anderson & Freebody, 1983; Cooksey & Freebody, 1987; Zimmerman, Broder, Shaughnessy, and 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 via alphabetization. A copy of this verbal measure is found in Appendix F. The subjects 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 have also been previously used (see Snodgrass & Corwin, 1988) resulting in virtually identical correlational results.

The nonverbal measure consisted of 18 problems from Raven's Advanced Progressive Matrices (Set II, Raven, 1962), a task tapping analytic intelligence and commonly viewed as a good measure of $g$ (Carpenter et al., 1990). The students were given 15 minutes to complete the 18 items on the test. By
Belief Bias

57

eliminating twelve of the easiest problems, where performance in a college sample is near ceiling (Carpenter et al., 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) an attempt was made to achieve a cut-time version of the Advanced Matrices that would still have adequate reliability and discriminating power. A previous investigation used a 16-item version of the Standard Progressive Matrices for cut-time administration and achieved reliabilities over .75 in samples of children (Cahan & Cohen, 1989). The GCA2 index was created by standardizing the raw scores on the vocabulary checklist measure and the Ravens Matrices and then adding the two standard scores together.

Thinking Dispositions Questionnaire

Participants completed a questionnaire consisting of a number of subscales. The response format for each item in the questionnaire was strongly agree (6), moderately agree (5) slightly agree (4), slightly disagree (3), moderately disagree (2), strongly disagree (1). The items from the subscales were randomly intermixed in the questionnaire. A brief description with some examples of these subscales follow:

Flexible Thinking Scale. Items on the Flexible Thinking Scale were devised by Stanovich and West for previous and ongoing research. The design of the items was influenced by a variety of sources from the critical thinking literature (e.g., Facione, 1992; Ennis, 1987; Lipman, 1991; Nickerson, 1987; Norris & Ennis, 1989; Perkins et al., 1993; Zechmeister & Johnson, 1992) but most specifically on the work of Baron (1985a, 1988, 1993) who has emphasized the concept of actively open-minded thinking through 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 ones beliefs in the face of contradictory evidence. There were ten items on
the Flexible Thinking Scale, some tapping the disposition toward reflectivity (e.g.,
"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,"
"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 (e.g., "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 (e.g., "There is
nothing wrong with being undecided about many issues," "Changing your mind
is a sign of weakness," "Basically, I know everything I need to know about the
important things in life,"--the latter two reverse scored).

**Counterfactual Thinking Scale.** A two-item scale designed to tap
counterfactual thinking was also devised by Stanovich and West (1997). 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 indicate
counterfactual thinking.

**Absolutism.** 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,
Belief Bias

1992; Schommer, 1990, 1993, 1994). Nine items designed to tap Perry's early stages which are characterized by an absolutist orientation were chosen for this scale. These early stages 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. 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”.

Dogmatism. The dogmatism subscale consisted of three items taken from a short-form field version (Troldahl, & Powell, 1965) of Rokeach's (1960) dogmatism scale. The three items were “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”.

Belief Identification. This scale was inspired by a theoretical paper from Cederblom (1989), in which he argues that a potential dimension of importance in explaining how willing people are to reason about issues involves how closely people identify their beliefs with their concept of “self”. At one extreme of this dimension, Cederblom argues, are people who closely identify their beliefs with their own sense of persona. These people may in consequence be less inclined to revise or change beliefs since they view them as such an intimate part of who they are. At the other end of the spectrum are those who see themselves simply as a belief forming-process. These people don't identify their beliefs with their sense of self and simply view belief as external entities that may be evaluated, revised or
discarded without impacting on their own persona. As a consequence of this favorable view, Cederblom argues, people would be less reluctant to depart with previously held beliefs and thus are less susceptible in recruiting biased strategies. Kuhn (1992) has also emphasized the importance of this sort of dimension on peoples' ability to engage in constructive argumentation. Referring to this dimension under the rubric of “belief ownership”, Kuhn makes the case that having a sense of “ownership” in your beliefs can be detrimental to a fair evaluation of their actual validity.

A nine item scale was distilled from the ideas expressed in Cederblom (1989). Examples of some of these items are “Someone who attacks my beliefs is not insulting me personally”--reverse scored, “It makes me happy and proud when someone famous holds the same beliefs that I do” and “What beliefs you hold have more to do with your own personal character than the experiences that may have given rise to them”.

Impulsiveness. The eight items from the Neuroticism-Impulsiveness facet of the Revised NEO Personality Inventory (Costa & McCrae, 1992) were administered (e.g., “I have trouble resisting my cravings”, “I rarely indulge in anything”--the latter reverse scored).

Deliberation. The eight items from the Conscientiousness-Deliberation facet of the Revised NEO Personality Inventory (Costa & McCrae, 1992) were also administered (e.g., “I plan ahead carefully when I go on a trip”, “I often do things on the spur of the moment”--the latter reverse scored).

Openness-Values. The eight items from the Openness-Values facet of the Revised NEO Personality Inventory were also 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).
Openness-Ideas. Also administered from the Revised NEO Personality Inventory (Costa & McCrae, 1992) were the eight items from the Openness-Ideas facet (e.g., “I have a lot of intellectual curiosity”, “I find philosophical arguments boring”–the latter reverse scored).

Categorical Thinking. Three items from the categorical thinking subscale of Epstein and Meier’s (1989) Constructive Thinking Inventory (CTI) 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”; “I tend to classify people as either for me or against me.”

Need for Closure. Twelve items taken from Kruglanski et al. (1993) Need for Closure Scale were administered. This scale purports to measure a cluster of related motivational tendencies such as a desire for predictability, preference for order and structure, discomfort with ambiguity and desire for definite knowledge on some issue. Kruglanski et al. refer to these motivational tendencies as the need for cognitive closure. Examples of some of the items used in this study include “I dislike it when a person’s statement could mean many different things”, “I always see many possible solutions to problems I face”–the latter reverse scored.

Faith in Intuition. Seven items were taken from a 12 item scale developed by Epstein, Pacini, Heier & Denes-Raj (1996). The Faith in Intuition Scale was designed by its authors to measure peoples’ engagement and confidence in their own intuitive abilities. Epstein et al. propose two modes of thinking style which they call the rational system and the experiential system. The Faith in Intuition scale was developed as an attempt to obtain individual difference measures of the experiential system. Examples of some of the items used in this study include “I believe in trusting my hunches”, and “I can usually feel when a person is right or wrong even if I can’t explain how I know”.

Head Over Heart. Eight items were taken from a scale developed by Epstein, Pacini, Heier & Denes-Raj (1996). Similar to the Faith in Intuition scale, this scale was designed by its authors to measure peoples general engagement and confidence in relying on an analytic mode of thinking over a more emotional based mode. Examples of some items used in this scale include “For important decisions, I tend to do the logical thing, no matter how I feel about the situation”, and “I believe in following my heart more than my head”—the latter reverse scored.

Paranormal Beliefs. The Paranormal beliefs subscale was composed of six items. Two items were concerned with belief in astrology (“It is advisable to consult your horoscope daily”; “Astrology can be useful in making personality judgments”) and were adapted from the paranormal belief scale validated by Jones, Russell, and Nickel (1977). The four remaining items concerned the belief in the concept of luck (e.g., “I have personal possessions that bring me luck at times”, “The number 13 is unlucky”) and were similar to items on the superstition subscale of a paranormal beliefs questionnaire developed by Tobacyk and Milford (1983).

Superstitious Thinking. Three items from the Superstitious Thinking subscale of Epstein and Meier’s Constructive Thinking Inventory (CTI) were administered: “I have found that talking about successes that I am looking forward to can keep them from happening”, “I do not believe in any superstitions”—reverse scored, and “When something good happens to me, I believe it is likely to be balanced by something bad”.

Consideration of Future Consequences Scale. Participants filled out this 12 item scale developed by Strathman, Gleichner, Boninger & Edwards (1994). The Consideration of Future Consequences Scale (CFC) was designed to measure individual differences in the extent to which people take into consideration distant
as opposed to the immediate consequences associated with potential courses of action. Examples of some of these items are “I am willing to sacrifice my immediate happiness or well-being in order to achieve future outcomes”, “I only act to satisfy immediate concerns, figuring that I will take care of future problems that may occur at a later date”—the latter reverse scored.

**Concept of Luck.** Four items reflecting conceptions of luck were taken from previous work by Stanovich and West (1998). Some examples of these items are “I have personal possessions that bring me luck at times” and “The number 13 is unlucky.”

**Response Bias and Effort Checks**

**Impression Management.** Five items reflecting social desirability response bias were taken from (Paulhus, 1991). Some examples of these items are “I always obey laws, even if I'm unlikely to get caught”, “I sometimes tell lies if I have to”—the latter reverse scored.

**Effort Probe.** One effort probe item was situated about two thirds through the questionnaire. This item read “As a scoring check, please mark strongly agree for this item.” The effort probe was used in concert with indices of task performance as a converging indication that a given participant was not exhibiting an honest effort in performing the tasks.

**Demographic and Print Exposure Measures**

Participants indicated some potentially relevant demographic information on a demographic form. Obtained was information relating to educational history, educational aspirations, parents educational background etc. Also administered were two measures of print exposure developed by Stanovich and Cunningham (1992). Both the Author Recognition Test (ART) and the Magazine Recognition Test (MRT) employed the same checklist-with-foils format as the Vocabulary Test described above. The stimuli used for the ART were 45 author
names and 41 nonexisting author names. The real author names and nonexistence author names were intermixed via alphabetization. The subjects were told that some of the names were of actual authors and that others were not and that their task was to read through the list of names and to put a check mark next to those that they knew to be actual authors. 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 Magazine Recognition Test (MRT) was structured in an identical fashion to the ART but with the exception of magazine titles rather than author names. The MRT consisted of 60 actual magazine titles and 34 nonexisting magazine titles. Again the scoring of this test was done in identical fashion to the ART. A copy of both the ART and MRT is found in Appendix G.

Procedure. Participants were all tested one at a time with the same experimenter. All participants read and signed a consent form prior to commencing the testing. The vast majority of participants encountered the common test battery in the same sequential order. Table 1 lists the order of task administration. Eight very brief vignettes requiring a judgment/prediction which are not part of this current study were interspersed in the test battery. The time to complete each of these vignettes ranges roughly from 1 minute to 3 or so minutes. The sequential placement of these vignettes are also indicated on Table 1. Furthermore, 117 of the participants completed an additional task not part of this reported study. This additional task was situated at the very end of the testing battery and required about 3 to 5 minutes to complete.

At about halfway through the battery participants were asked by the experimenter if they would like to take a break. Sixteen participants (12.7%) took a break. All breaks, with one exception, were about 5 minutes in duration. One participant required about 20 minutes for religious prayer. Prior to beginning,
participants had also been informed that during the testing they could break any
time they wish. Only two of these 16 participants had asked to take a break, one
prior to the scheduled break the other after the scheduled break. All participants
were also offered a soft drink and a snack at the scheduled break regardless of
whether they chose to break or not.

Table 1

Order of Task Administration

<table>
<thead>
<tr>
<th>Serial position</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Demographics Form</td>
</tr>
<tr>
<td>2.</td>
<td>Vignette 1*</td>
</tr>
<tr>
<td>3.</td>
<td>Argument Evaluation Test I  (Prior Belief)</td>
</tr>
<tr>
<td>4.</td>
<td>Vignette 2*</td>
</tr>
<tr>
<td>5.</td>
<td>Dispositions Questionnaire</td>
</tr>
<tr>
<td>6.</td>
<td>Vignette 3*</td>
</tr>
<tr>
<td>7.</td>
<td>Syllogisms</td>
</tr>
<tr>
<td>8.</td>
<td>Vignette 4*</td>
</tr>
<tr>
<td>9.</td>
<td>Confounded Variable Task†</td>
</tr>
<tr>
<td>10.</td>
<td>Vignette 5*</td>
</tr>
<tr>
<td>11.</td>
<td>Ecological Height Judgment Task</td>
</tr>
<tr>
<td>12.</td>
<td>Matched Height Judgment Task</td>
</tr>
<tr>
<td>13.</td>
<td>Opinion Prediction I (Other Prediction)</td>
</tr>
<tr>
<td>14.</td>
<td>Knowledge Test &amp; Other Prediction†</td>
</tr>
<tr>
<td>15.</td>
<td>WAIS-R Vocabulary subtest</td>
</tr>
<tr>
<td>16.</td>
<td>WAIS-R Block Design subtest</td>
</tr>
<tr>
<td>17.</td>
<td>Break (5 minutes) -- Optional</td>
</tr>
<tr>
<td>18.</td>
<td>Argument Evaluation Test II (evaluation)</td>
</tr>
<tr>
<td>19.</td>
<td>Covariation Judgment I (Prior Belief)</td>
</tr>
<tr>
<td>20.</td>
<td>Vignette 6*</td>
</tr>
<tr>
<td>21.</td>
<td>Opinion Prediction II (Agreement)</td>
</tr>
<tr>
<td>22.</td>
<td>Vignette 7*</td>
</tr>
<tr>
<td>23.</td>
<td>Vignette 8*</td>
</tr>
<tr>
<td>24.</td>
<td>Additional Task not part of present study*</td>
</tr>
</tbody>
</table>

* not part of this study
† Non central tasks which may be bumped to end of battery in the event of time problems

The average time required to complete the test battery was 182.9 minutes (SD = 28.9 minutes). Because there was enormous variation in the time participants
took to complete the tasks, task order was modified for a few participants to
ensure that the tasks of most relevance to this study were completed.

Experimental tasks that were deemed less central to this study were the
knowledge prediction task (KPT), and the confounded variable task (CVT).
Problems with time were dealt by bumping the KPT and CVT to the end of the
testing battery. These few order modifications were made to resolve one of 2
potential logistical problems which arose during a testing session: 1) at some point in the testing the participant expressed some need or desire to be done at a stated time which the experimenter estimated may not be attainable given the participants pace, or 2) the experimenter estimated that the participant in question would run significantly longer than the 3 hour 15 minute estimate and importantly, create further complications such as significantly delaying the testing of the next scheduled participant. The latter situation occurred twice and resulted in bumping both the CVT and KPT to the end of the battery. In both cases neither participant completed those two tasks. The former situation was more frequent and resulted in bumping the CVT and KPT to the end for 11 participants. One participant still managed to complete all the tasks regardless of the concern with time, 5 participants did not complete the CVT, 3 participants did not complete the KPT and 2 participants did not complete both of these less central tasks. One participant was extremely slow in working through the tasks, and the task order modification was not able to salvage the completion of all the tasks deemed central. This participant did not complete the noncentral CVT and KPT as well as the more important opinion prediction task, and covariation judgment task.

The knowledge prediction task was also not completed by 4 other participants for the following reason. During the study some participants who spoke English as a second language stated that although they did know the answer to some KPT items and could answer them in their first language, they could not recall the English word/name for the item. This turned out to be indigenously problematic to the very nature of the task. At the time of testing it was decided that these four participants (all having English as a second language with an appreciable accent) would not be administered the KPT. A successful effort was made to ensure that all future scheduled participants displayed no appreciable accent.
Table 2 summarizes this breakdown of tasks not completed by the participants.

Table 2

Summary of Tasks in the Testing Battery That Were Not Completed in Proper by Participants.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Confounded Variable Task</th>
<th>Knowledge Prediction Task</th>
<th>Height Judgment Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>#22</td>
<td>N/C (time)</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#26</td>
<td>-</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#35</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#47</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#53</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#60</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#61</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#62</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#64</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#65</td>
<td>-</td>
<td>N/C (Eng. 2nd)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#68+</td>
<td>N/C (time)</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#74</td>
<td>-</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#82</td>
<td>N/C (time)</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#83</td>
<td>N/C (time)</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#84</td>
<td>N/C (time)</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#89</td>
<td>N/C (time)</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#102</td>
<td>N/C (time)</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#109</td>
<td>-</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#112</td>
<td>N/C (time)</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
<tr>
<td>#126</td>
<td>-</td>
<td>N/C (time)</td>
<td>Outlier</td>
</tr>
</tbody>
</table>

Note: N/C = Not Complete, time = time problem issue; Eng. 2nd = English 2nd Language issue; Outlier = Multivariate outliers having Mahalanobis distance metrics > 4.0; + = participant also did not complete covariation judgment and opinion prediction tasks.

In addition to these incomplete tasks, 7 participants were removed from the height judgment tasks in the ensuing analyses. These participants were multivariate outliers having Mahalanobis distance metrics greater than 4.0 (see Tabachnick & Fidell, 1983). In total, 124 participants completed the syllogism test, the AET and all the cognitive ability, thinking disposition, print exposure and demographic measures. A total 123 participants completed the covariation judgment task and the opinion prediction task, 117 participants completed the height judgment tasks in proper, 113 participants completed the confounded variable task, and 115 participants completed the knowledge prediction task.
Overview of the Methodological Approach

Figure 2 provides a concise visual schematic of the current methodological approach along with the relationships of interest to the present study. All participants will be administered a common test battery. Participants will be given six tasks that are designed to assess the extent of prior belief influence on reasoning and judgment (top right of Figure 2). The covariance structure between these prior belief tasks is of particular interest for various reasons outlined above. Participants will also complete two social perception tasks (one drawn from and the other inspired by current tasks existing in the social perception literature). The covariance structure of these social perception tasks with one another, as well as with the prior belief tasks is of interest to this investigation. Like the prior belief tasks, the social perception tasks can be seen as reasoning tasks in which people are found to project beliefs onto the world. It is for this reason that these social perception tasks will largely be referred to simply as projection tasks. Whereas prior belief influences on reasoning and judgment is the form of belief projection of concern in the prior belief tasks, belief projection in the social perception tasks manifests itself in the form of projecting consensus on the prediction of opinions or the knowledge of others. Also of importance is accounting for variation on these tasks using a cognitive capacity-thinking disposition framework (see the left side of Figure 2). To accomplish this, participants will also complete measures of cognitive capacity and thinking dispositions. The consequences of individual differences in both the cognitive capacity measures and thinking disposition measures on task performance will be extensively examined. Data was also collected on certain demographic variables and reading exposure measures.
Figure 2. Visual schematic of methodological approach and relationships of interest. Note: * = tasks using ecologically valid stimuli.
Results and Discussion

Individual Differences in
Prior Belief Task Performance

Participants were found to vary substantially on the performance indices of all the reasoning tasks. The six Panels of Figure 3 display the distributions of some of the focal indices of performance for all the prior belief reasoning and judgment tasks. Panel A of Figure 3 is the belief bias index score for syllogistic reasoning. Recall that this score was obtained by subtracting the inconsistent syllogism score from the consistent syllogism score. The mean syllogistic belief bias scores ranged from a -1 to +8 with most of participants falling between 0 (no belief bias) and +3 (moderate belief bias). The mean belief bias score was 1.53 (SD = 1.77). As can be seen in Panel A of Figure 3, this distribution was skewed with the modal score being zero and a sharp drop off in frequency for scores below zero.

Panel B of Figure 3 is the belief congruency score for the confounded variable task. Recall that this index is designed such that participants indicating lower values on this score signal that they are correctly engaging in a more complex ANCOVA-like reasoning and were thus able to discern that the true relation amongst the variables is opposite to what they had previously indicated in their opinion. Higher scores on this index signal a failure to use the more complex ANCOVA-like strategy by defaulting to a simplistic aggregate analysis leading to a conclusion congruent with the stated prior belief. Scores on this index ranged from a low of 2 to a high of 9. The mean belief congruence score was found to be 6.29 (SD = 1.54). Once again as can be seen in Panel B of Figure 3, this distribution was somewhat skewed with the modal belief congruency score being 7.

Panel C and Panel D of Figure 3 outline an identical index used for the
ecological and matched height judgment tasks respectively. This index is simply the point biserial correlation between the gender cue (G, scored 0/1) and the ensuing height estimates (E), i.e., r(G,E). There are two notable differences in

![Figure 3](image-url)
this index between the ecological and matched tasks. The ecological response
distribution is both less variable and centered at a notably higher value. Whereas
the average ecological task correlation between the Gender cue and and ensuing
estimates, i.e. r(G,E), in a participant’s responses was found to be 0.693 (SD =
.129; range from low of .361 to high of .953), the average of this same index on the
matched task was notably lower at 0.281 (SD = .196; range from low of -.189 to high
of .865). This difference was highly significant, t(116) = 26.37, p < .0001. Thus in
the aggregate, participants proceeding from the ecological task to the matched
task do decrease their reliance on the gender cue. The extent to which they do
this, however, is remarkably variable.

Panel E and Panel F of Figure 3 also outline identical performance indices
used in the analysis of the argument evaluation test (AET) and covariation task
respectively. On the AET, the mean multiple correlation across the 124
regressions in which each participant’s evaluations were regressed on the
objective argument quality scores and on his or her prior opinion scores was .534
(SD = .159). The mean standardized beta weight for argument quality as
displayed in Panel E was found to be 0.413 (SD = .211; range from a low of -.193 to a
high of .834). On the covariation task, the mean multiple correlation across the
123 regressions in which each participant’s evaluations were regressed on the Δp
statistic and on his or her prior opinion scores was .693 (SD = .208). The mean
standardized beta weight for the Δp statistic as displayed in Panel F was found to
be 0.587 (SD = .243; range from a low of -.135 to a high of .967). Although
performance of several of these tasks can be indexed in slightly different ways,
those outlined above provide a good illustration for the large variation in
individual differences of performance across all the prior belief reasoning and
thinking tasks.

The Correlates of Normative
Responses on the Prior Belief Tasks
In an attempt to address the nature of the differences between participants who exhibit the normative versus nonnormative responses on the prior belief tasks, median splits on a focal performance index for each task was performed and the resulting two groups were compared on a series of variables that would seem to have relevance for characterizing the two resulting groups. These variables of relevance are the two cognitive capacity measures (general cognitive ability 1 and general cognitive ability 2), high school and university averages as reported by participants, scores on the two indices of print exposure viz., the author recognition test (ART) and the magazine recognition test (MRT), and finally the score on the general knowledge test.*

A median split was performed on a focal index of each task such that the resulting series of two groups can be viewed as the nonnormative-response group (LO-NORM) and the normative-response group (HI-NORM). Table 3 displays the results of a series of t-tests between these groups alternating the above outlined variables as the criterion measures. Participants on the syllogisms task were split on the belief bias index. The mean of the resulting high belief bias (LO-NORM) and low belief bias (HI-NORM) groups was found to be 2.79 (SD = 1.68) and 0.27 (SD = 0.55) respectively. These two resulting groups differed significantly on the belief bias index t(122) = -11.21, p < .0001. As can be seen in Table 3, the HI-NORM group was also found to score significantly higher on both measures of cognitive capacity and report having obtained significantly higher high school and university averages than the LO-NORM group.

A median split was also performed on the beta weight for argument quality on the argument evaluation test. The mean beta weight for the low argument quality beta weight group (LO-NORM) and high argument quality beta weight group (HI-NORM) were found to be 0.248 (SD = 0.15) and 0.578 (SD = .108) respectively. Once again this difference between the LO-NORM and HI-NORM

* The mean score on the general knowledge test was found to be 14.59 (SD = 4.92).
Table 3
Mean Scores of Participants with the Lowest and Highest Scores on the Normative Performance Index for the Syllogisms Test, Argument Evaluation Test, Covariation Judgment Task, and Confounded Variable Task (Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th>Task Variable</th>
<th>N</th>
<th>LO-NORM</th>
<th>HI-NORM</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syllogisms</strong></td>
<td>124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td></td>
<td>104.40 (11.44)</td>
<td>112.29 (10.63)</td>
<td>3.98****</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td></td>
<td>-0.60 (1.40)</td>
<td>0.60 (1.35)</td>
<td>4.88****</td>
</tr>
<tr>
<td>High School Avg.</td>
<td></td>
<td>83.21 (5.79)</td>
<td>85.60 (6.49)</td>
<td>2.10*</td>
</tr>
<tr>
<td>University Avg.</td>
<td></td>
<td>71.55 (7.22)</td>
<td>75.36 (5.99)</td>
<td>2.91**</td>
</tr>
<tr>
<td>Author Recognition</td>
<td></td>
<td>0.42 (0.20)</td>
<td>0.48 (0.17)</td>
<td>1.93</td>
</tr>
<tr>
<td>Magazine Recognition</td>
<td></td>
<td>0.53 (0.175)</td>
<td>0.54 (0.14)</td>
<td>0.39</td>
</tr>
<tr>
<td>General Knowledge</td>
<td></td>
<td>13.93 (4.54)</td>
<td>15.24 (5.22)</td>
<td>1.44</td>
</tr>
<tr>
<td><strong>Argument Evaluation</strong></td>
<td>124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td></td>
<td>104.50 (12.23)</td>
<td>112.19 (9.80)</td>
<td>3.87***</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td></td>
<td>-0.41 (1.38)</td>
<td>0.41 (1.51)</td>
<td>3.19**</td>
</tr>
<tr>
<td>High School Avg.</td>
<td></td>
<td>83.98 (6.19)</td>
<td>84.85 (6.32)</td>
<td>0.75</td>
</tr>
<tr>
<td>University Avg.</td>
<td></td>
<td>71.77 (6.67)</td>
<td>75.36 (6.61)</td>
<td>2.73**</td>
</tr>
<tr>
<td>Author Recognition</td>
<td></td>
<td>0.41 (0.19)</td>
<td>0.49 (0.18)</td>
<td>2.24*</td>
</tr>
<tr>
<td>Magazine Recognition</td>
<td></td>
<td>0.51 (0.16)</td>
<td>0.56 (0.15)</td>
<td>1.79</td>
</tr>
<tr>
<td>General Knowledge</td>
<td></td>
<td>13.41 (5.40)</td>
<td>15.64 (4.23)</td>
<td>2.48*</td>
</tr>
<tr>
<td><strong>Covariation Judgment</strong></td>
<td>123</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td></td>
<td>102.81 (10.45)</td>
<td>114.16 (10.04)</td>
<td>6.15****</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td></td>
<td>-0.71 (1.30)</td>
<td>0.70 (1.36)</td>
<td>5.88****</td>
</tr>
<tr>
<td>High School Avg.</td>
<td></td>
<td>84.15 (6.28)</td>
<td>84.72 (6.25)</td>
<td>0.49</td>
</tr>
<tr>
<td>University Avg.</td>
<td></td>
<td>72.28 (7.33)</td>
<td>74.73 (6.19)</td>
<td>1.83</td>
</tr>
<tr>
<td>Author Recognition</td>
<td></td>
<td>0.41 (0.17)</td>
<td>0.50 (0.19)</td>
<td>2.75**</td>
</tr>
<tr>
<td>Magazine Recognition</td>
<td></td>
<td>0.50 (0.15)</td>
<td>0.58 (0.16)</td>
<td>2.82**</td>
</tr>
<tr>
<td>General Knowledge</td>
<td></td>
<td>12.95 (4.35)</td>
<td>16.32 (4.98)</td>
<td>3.90***</td>
</tr>
<tr>
<td><strong>Confounded Variable</strong></td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td></td>
<td>105.95 (12.53)</td>
<td>111.46 (10.32)</td>
<td>2.55*</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td></td>
<td>-0.50 (1.49)</td>
<td>0.56 (1.33)</td>
<td>3.99****</td>
</tr>
<tr>
<td>High School Avg.</td>
<td></td>
<td>84.32 (5.43)</td>
<td>85.23 (6.24)</td>
<td>0.81</td>
</tr>
<tr>
<td>University Avg.</td>
<td></td>
<td>73.77 (6.32)</td>
<td>73.33 (7.84)</td>
<td>0.29</td>
</tr>
<tr>
<td>Author Recognition</td>
<td></td>
<td>0.39 (0.17)</td>
<td>0.52 (0.19)</td>
<td>3.95****</td>
</tr>
<tr>
<td>Magazine Recognition</td>
<td></td>
<td>0.48 (0.16)</td>
<td>0.61 (0.13)</td>
<td>4.51****</td>
</tr>
<tr>
<td>General Knowledge</td>
<td></td>
<td>13.59 (4.34)</td>
<td>15.83 (5.35)</td>
<td>2.39*</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001, all two-tailed.
Note: All above Ns lose at most an additional 7 participants in High School Avg. variable (scores not reported), 22 participants in the University Avg. variable (scores not reported), and 9 participants in the General Knowledge variable (test not administered).
groups was found to be highly significant, $t(122) = -13.97, p < .0001$. Table 3 shows that the HI-NORM group also scored significantly higher on both cognitive capacity measures, reported university average, the author recognition test (ART), and the general knowledge test.

Presented next is the median split analysis based on the partitioning of the $\Delta p$ statistic from the covariation judgment task. The mean for the group with relatively low beta weight values for the $\Delta p$ statistic (LO-NORM) and the group with relatively high beta weight values for the $\Delta p$ statistic (HI-NORM) was found to be 0.399 (SD = .194) and 0.779 (SD = .086) respectively. These two groups were also significantly different from one another $t(121) = -14.01, p < .0001$. Again the HI-NORM group scored significantly higher on both cognitive capacity measures, both indices of print exposure (ART and MRT), and the general knowledge test.

Finally, the confounded variable task was also subjected to the same analysis by using the belief congruency score as the variable for the median split. The resulting LO-NORM group (high belief congruency score through the use of a simple aggregate approach) had a mean score of 7.44 (SD = .66) and the HI-NORM group (low belief congruency score through the use of an ANCOVA-like approach) had a mean score of 5.13 (SD = 1.28). These two groups were furthermore found to be significantly different from one another $t(111) = 12.13, p < .0001$. It was also the case that the HI-NORM group in this task scored significantly higher in both cognitive capacity measures as well as both print exposure measures (ART and MRT) and the general knowledge test.

Thus in all four of these prior belief tasks displayed in Table 3, participants falling in the HI-NORM groups consistently scored significantly higher in the cognitive capacity measures, they were also more likely to have achieved higher high school and university averages, experienced more exposure to printed materials, and were also found to be generally more knowledgeable than those
participants falling in the LO-NORM groups.

These prior belief tasks were designed such that projection of belief would lead to nonnormative performance. But in the real world, projection of belief need not necessarily lead to inaccuracy in judgment or reasoning. This is true since projection of accurate beliefs may facilitate the efficacy of reasoning. Thus we would not expect that those who do project belief as necessarily less cognitively competent. The analysis summarized in Table 4 reinforces this point. Table 4 is identical in structure to Table 3, except that it focuses on the height judgment tasks. Once again median splits are performed on various indices of belief projection without assigning the normative status as was done in Table 3. Recall that the r(G,E) index is calculated for each individual participant as the point biserial correlation between the gender cue and the participant’s ensuing height estimates. The higher this index, the larger the extent to which the participant is projecting the prior belief about the relationship between gender and height.

Of particular note in Table 4 is the pattern of differences in the variable scores between those exhibiting high levels of belief projection (HI-projection) in the ecological height judgment task and all those groupings of high belief projection observed in Table 3. Whereas groupings of participants who do not project belief and who successfully reason independently of belief in the tasks displayed in Table 3 score higher on the various variables displayed, it is those participants who do project belief in the ecological height judgment task, i.e., high correlations between the gender cue and height estimates, who score significantly higher on both cognitive capacity measures, and approach significantly higher scores on reported high school average and both the print exposure variables. This is the opposite pattern to that observed in Table 3.

In addition to this, the Table 3 pattern (i.e., unfavourable performance on the displayed variables for those projecting belief--the LO-NORM group) emerges
once again in Table 4 for indices of performance on the matched height judgment task. Recall that in the matched height judgment task, the gender cue is rendered nondiagnostic and participants are made fully aware of this fact. Thus

Table 4

Mean Scores of Participants with the Lowest and Highest Scores on the Normative Performance Index for the Syllogisms Test, Argument Evaluation Test, Covariation Judgment Task, and Confounded Variable Task (Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th>Task Variable</th>
<th>LO-Projection</th>
<th>HI-Projection</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecological r(G,E)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td>106.47 (11.29)</td>
<td>111.64 (10.99)</td>
<td>2.51*</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td>-0.23 (1.48)</td>
<td>0.37 (1.45)</td>
<td>2.22*</td>
</tr>
<tr>
<td>High School Avg.</td>
<td>83.58 (6.29)</td>
<td>85.64 (5.85)</td>
<td>1.78†</td>
</tr>
<tr>
<td>University Avg.</td>
<td>73.31 (7.28)</td>
<td>73.82 (6.69)</td>
<td>0.36</td>
</tr>
<tr>
<td>Author Recognition</td>
<td>0.43 (0.19)</td>
<td>0.49 (0.17)</td>
<td>1.82†</td>
</tr>
<tr>
<td>Magazine Recognition</td>
<td>0.53 (0.16)</td>
<td>0.57 (0.13)</td>
<td>1.79†</td>
</tr>
<tr>
<td>General Knowledge</td>
<td>14.69 (5.34)</td>
<td>14.59 (4.34)</td>
<td>0.12</td>
</tr>
<tr>
<td>Matched r(G,E)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td>110.92 (11.85)</td>
<td>107.27 (10.72)</td>
<td>1.74†</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td>0.30 (1.59)</td>
<td>-0.16 (1.36)</td>
<td>1.69†</td>
</tr>
<tr>
<td>High School Avg.</td>
<td>83.98 (6.19)</td>
<td>85.32 (6.05)</td>
<td>1.15</td>
</tr>
<tr>
<td>University Avg.</td>
<td>75.25 (5.46)</td>
<td>71.85 (7.90)</td>
<td>2.46*</td>
</tr>
<tr>
<td>Author Recognition</td>
<td>0.47 (0.19)</td>
<td>0.45 (0.18)</td>
<td>0.47</td>
</tr>
<tr>
<td>Magazine Recognition</td>
<td>0.56 (0.15)</td>
<td>0.54 (0.15)</td>
<td>0.61</td>
</tr>
<tr>
<td>General Knowledge</td>
<td>15.40 (5.19)</td>
<td>13.93 (4.37)</td>
<td>1.61</td>
</tr>
<tr>
<td>HJ Adjustment Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td>113.22 (11.27)</td>
<td>104.86 (9.96)</td>
<td>4.25****</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td>0.76 (1.36)</td>
<td>-0.63 (1.28)</td>
<td>5.71****</td>
</tr>
<tr>
<td>High School Avg.</td>
<td>84.89 (6.04)</td>
<td>84.43 (6.26)</td>
<td>0.39</td>
</tr>
<tr>
<td>University Avg.</td>
<td>74.04 (7.59)</td>
<td>73.08 (6.29)</td>
<td>0.68</td>
</tr>
<tr>
<td>Author Recognition</td>
<td>0.51 (0.17)</td>
<td>0.40 (0.18)</td>
<td>3.54***</td>
</tr>
<tr>
<td>Magazine Recognition</td>
<td>0.59 (0.13)</td>
<td>0.51 (0.16)</td>
<td>3.16**</td>
</tr>
<tr>
<td>General Knowledge</td>
<td>15.57 (5.34)</td>
<td>13.67 (4.04)</td>
<td>2.11*</td>
</tr>
</tbody>
</table>

† = p < .10, * = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001, all two-tailed.

accurate height judgments in the matched task demands that the gender cue is
not utilized as in the ecological task, i.e., ideally \( r(G,E) \) should equal zero. As shown in Panel C and D of Figure 3, although in the aggregate participants did make less use of this gender cue in the matched task as opposed to the ecological task, there was much variability in the extent to which participants inhibited the gender cue influence on height estimation.

A median split on the \( r(G,E) \) index for the matched task is displayed next in Table 4. Although failing to reach statistical significance, the HI-Projection Group (as was the case in the tasks displayed in Table 3) was found to score lower on measures of cognitive capacity, and did attain significantly lower scores on reported high school average than the LO-Projection group. The projection of belief in the matched task can also be indexed with the HJ Adjustment score. As outlined earlier, the HJ Adjustment score takes into account the magnitude of the belief adjustment required to bring a invalidated prior belief into congruence with the task environment. Performing a median split on this index convincingly displays that the HI-Projection of belief group in the matched task was significantly lower on both measures of cognitive ability, significantly lower on both print exposure measures, and achieved a significantly lower score on the general knowledge test than did LO-projection group.

**Consequences of Projecting Belief on the Accuracy of Tracking the Environment**

The preceding analysis has shown that participants having higher levels of cognitive ability were more likely to project belief in the ecological height judgment task and less likely to project belief in the matched height judgment task. What are the resulting consequences for height judgment accuracy? Height judgment accuracy can be indexed using at least two performance indices. First we can look at the correlation between the height estimates and the actual target heights i.e., \( r(T,E) \). This index alone is not sufficient since it will only indicate the accuracy in ordering the magnitude of height estimate correctly
with the magnitude of the actual target heights. What it will not reveal is the correct use of the scale (see Brown & Siegler, 1993). Thus, for example, a participant who succeeds in ordering the magnitude of their height estimates in accord with the order of the actual target heights will attain a perfect correlation of 1.0 between target height and height estimate i.e., \( r(T,E) \), despite the fact that they may have consistently overestimated the targets to be 3 inches taller than they really were. In order to capture this additional aspect of proper scale use, a second index of performance is derived for each participant by taking the sum of the absolute deviations between their height estimate and actual target height (SAD) across all the stimuli. Thus SAD will provide a means to index the extent to which estimates actually match the target heights.

Participants' estimates in the ecological task were highly correlated with the gender cue viz., \( r(G,E) \). This fact proves to be efficacious for the resulting height judgment accuracy. This is true since there actually is a correlation between target gender and target height in the ecological set (\( r(G,T) = .628 \)). The efficacy of projecting the gender/height relationship is seen in the significant correlation between the \( r(G,E) \) measure of projection of belief with both the \( r(T,E) \) accuracy measure (\( r = .697, p < .001 \)) and the SAD accuracy measure (\( r = -.525, p < .001 \)). Thus in terms of both properly ordering the stimuli in height magnitude and the absolute deviations of target height estimates from the actual target heights, participants projecting a prior belief about the height/gender relationship in the ecological height judgment task gave more accurate height estimates.

In their original study, Nelson et al. (1990) reported that their participants had overused the gender cue. Male targets, they reported, were estimated to be on average taller than female targets even after statistically controlling the actual difference between males and females. Nelson et al. (1990) did not, however, investigate the consequences of this “overuse” of the gender cue on the accuracy of
height judgment. As previously mentioned, this is the exact issue that has surfaced in the literature on the so-called false consensus effect. As for projecting the gender/height relationship in the ecological height judgment task, the present results suggest that even this apparent "overuse" of the gender cue (viz., the application of a stereotype) was efficacious. The reason for this can be best explained in reference to the top half of Figure 4. The top half of Figure 4 diagrams the outcome when the aggregate height estimates (averaged over the 117 participants in the ecological task) are regressed on both the height estimate predictors of target gender and actual height. The two resulting standardized beta weights for gender and actual height are also given in the top half of the diagram (the ecological environment portion). The standardized beta weights give an indication of the influence of each predictor on the height judgments with the other predictor statistically controlled. Gender as a predictor of height judgment attained a highly significant beta weight of 0.610, \( p < .001 \). Once more, this diagram also shows that gender in the ecological task is a significant predictor of actual height (standardized beta weight = 0.628, \( p < .001 \)).

The same regression analysis displayed in this diagram for the aggregate responses was repeated for each of the 117 participants who completed this task in proper. Each participant had his or her height estimates simultaneously regressed on both the two predictors of target gender and actual target height. This resulted in sets of standardized beta weights for the gender cue and actual height. The gender beta weight (viz., an index for the influence of the gender cue after statistically controlling for actual height differences) was found to be positively correlated with \( r(T,E) \) (\( r = .443, p < .001 \)) and negatively correlated with SAD (\( r = -.314, p < .001 \)). Thus participants who overused the gender cue by judging males to be taller even after the actual difference in male and female target height was statistically controlled, were more accurate in both the \( r(T,E) \)
Belief Bias

and SAD indices. Nelson et al. (1990) neglected to address this important consequence of the observed projection of prior belief in this ecologically valid height judgment task. While participants did indeed "overuse" or over project the gender cue, prior belief dominance in the participants' height estimates actually facilitated their accuracy.

![Path diagram predicting the aggregate height judgments for both the ecological and matched tasks.](image)

**Figure 4.** Path diagram predicting the aggregate height judgments for both the ecological and matched tasks.

Note: Prediction of the aggregate height judgments in the ecological set displayed in top portion and prediction of the aggregate height judgments in the matched set displayed in bottom portion. Numbers on arrows are standardized beta weights.

* = p < .01, ** = p < .001

The consequence of projecting the gender cue in the matched task stands in stark contrast to doing so in the ecological task. The reason for this can best be explained by reference to the bottom half of Figure 4 (viz., the matched environment portion). The bottom half of Figure 4 diagrams the outcome when the aggregate height estimates (averaged over the 117 participants) in the
matched height judgment task are regressed on both the same height estimate predictors of target gender and actual target height. Although the beta weight for the gender cue in the matched height judgment condition is not as large as in the ecological condition [0.610, \( p < .001 \)], it is nevertheless statistically significant, [0.388, \( p < .05 \)]. Despite the extensive instructions to participants that they should ignore the gender cue in the matched task, as a group, participants' estimates were nevertheless significantly influenced by the target gender. Of particular importance here is the fact that gender is no longer a diagnostic cue for actual height in the matched condition, beta = .018, ns. It is for this reason that projecting the gender cue will hamper accurate judgments. Unlike the ecological task environment, the \( r(G,E) \) index in the matched task did not correlate with either \( r(T,E) \) (\( r = -.172, \) ns) or SAD (\( r = .143, \) ns). In fact the correlation signs on these two indices indicate that the more participants projected the gender/height relationship, the less accurate they were. An identical individual difference analysis to that performed on the ecological task also bears this out. Once again each participant's height judgments on the matched task were entered into a regression model with target gender and actual height as the two predictors. This time the resulting gender beta weight parameter failed to correlate negatively with SAD (\( r = .148, \) ns) and did correlate negatively with \( r(T,E) \) (\( r = -.181, \) \( p < .05 \))--i.e., rather than facilitating accuracy, the increased use of the gender cue resulted in less accuracy in the tendency to order the magnitude of the height estimates congruently with the actual heights.

**Exploring the Domain Specificity and Generality of Belief Bias**

The issue of domain specificity and generality is addressed in Table 5. As can be seen from Table 5, projecting the invalidated gender cue in the matched height judgment task, i.e, \( r(G,E)_{\text{matched}} \), was positively correlated with a direct index of belief projection viz., the belief bias index in syllogisms (belief bias score),
and negatively correlated with both successfully reasoning independently of prior belief in the AET (i.e., the beta weight for argument quality), and ability to properly track—in the face of conflicting beliefs—the actual bivariate correlation (i.e., the Δp statistic) existing between two variables in the covariation task. This was not the case for the projection cue in the ecological task, i.e., \( r(G,E)_{\text{ecological}} \). Whereas belief bias on syllogisms was positively correlated with \( r(G,E)_{\text{matched}} \), it was not correlated with \( r(G,E)_{\text{ecological}} \). Not only did \( r(G,E)_{\text{ecological}} \) fail to reach a significant correlation with the belief bias index for the syllogisms test, as \( r(G,E)_{\text{matched}} \) did, unlike the latter, it obtained a slight negative correlation with syllogistic belief bias \((r = -0.095)\) which was furthermore found to be significantly different than the correlation between \( r(G,E)_{\text{matched}} \) and belief bias \((t(115) = 3.52, p < .001; \text{test for difference in dependent correlations, Cohen & Cohen, 1983, p. 53})\). That \( r(G,E)_{\text{matched}} \) and \( r(G,E)_{\text{ecological}} \) obtain different correlation signs with the other prior belief tasks can be seen in Table 5 as you go down the two first columns. As in the case with the syllogistic belief bias index, \( r(G,E)_{\text{ecological}} \) differed from \( r(G,E)_{\text{matched}} \) in that it also failed to reach a significant correlation with the other prior belief task performance indices. Furthermore, with the exception of the confounded variable task, the correlation difference between \( r(G,E)_{\text{matched}} \) index and \( r(G,E)_{\text{ecological}} \) index with both of the remaining prior belief task indices in Table 5 were found to be significantly different from each other: AET (argument quality beta weight) \( t(115) = -4.30, p < .001; \) covariation task (Δp statistic) \( t(114) = -4.34, p < .001 \). Thus, it is the condition in which participants are required to inhibit projection of belief in a perceptual judgment task that is significantly positively correlated with belief bias on a logical reasoning task, negatively correlated with the ability to reason well independent of belief in an argument evaluation task, and negatively correlated with the ability to detect covariations between bivariate relations in the face of conflicting prior
beliefs.

Table 5
**Intercorrelations Among the Primary Prior Belief Task Variables.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (r(G,E), ) Matched</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. (r(G,E), ) Ecological</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. HJ Adjustment Score</td>
<td>-0.396***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Syllogisms, Belief Bias</td>
<td>0.209*</td>
<td>-0.095</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. AET</td>
<td>-0.261**</td>
<td>0.101</td>
<td>0.284**</td>
<td>-0.236*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Covariation Judgment</td>
<td>-0.200*</td>
<td>0.168</td>
<td>0.322***</td>
<td>-0.342***</td>
<td>0.324***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Confounded Variable</td>
<td>0.036</td>
<td>-0.046</td>
<td>-0.109</td>
<td>0.123</td>
<td>-0.188*</td>
<td>-0.127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Cognitive Ability 1</td>
<td>-0.155</td>
<td>0.256**</td>
<td>0.347***</td>
<td>-0.445***</td>
<td>0.449***</td>
<td>0.493***</td>
<td>-0.224*</td>
<td></td>
</tr>
<tr>
<td>9. Cognitive Ability 2</td>
<td>-0.215*</td>
<td>0.263**</td>
<td>0.461***</td>
<td>-0.495***</td>
<td>0.392***</td>
<td>0.505***</td>
<td>-0.329***</td>
<td>0.707***</td>
</tr>
</tbody>
</table>

* = \(p < .05\), ** = \(p < .01\), *** = \(p < .001\), all two tailed.

Parallel results were obtained when the HJ Adjustment score was examined. The HJ Adjustment score obtained a significant negative correlation with the belief bias score, and significantly positive correlations with the AET argument quality beta weight and covariation judgment \(\Delta p\) beta weight.

Performance on the syllogisms test, AET, and covariation judgment task correlated with one another in the expected way. Table 5 clearly shows that the belief bias score on the syllogisms test obtained significant negative correlations with the ability to reason independently of belief in both the AET and the covariation judgment task. Furthermore, the indices of reasoning well independent of prior belief on the AET and covariation judgment task were also significantly correlated in the expected direction (viz., a significant positive correlation).
These results are also of relevance to the normative issue since all but one of these tasks correlate with each other in a manner suggesting that participants who give the normative response in one task are also likely to give the normative response in another task. Furthermore, the normative response in tasks where the normative status might be questionable (e.g., the AET) is found to correlate with normative responses in tasks in were the normative status is not in question (e.g., syllogisms). Also relevant to the normative issue are the correlations involving the cognitive ability measures. These correlations converged with the relationships displayed in Tables 3 and 4. The cognitive ability measures were negatively correlated with the projection of belief on all but the ecological height judgment task—in which they were positively correlated with the projection of belief. Thus the correlation of projection of belief with cognitive ability was negative when projection was inefficacious and positive when projection was efficacious.

Both the domain specificity/generality issue and the relationship between cognitive ability and task performance was explored in a further series of hierarchical regression analyses. Table 6 presents the results of an hierarchical regression analysis in which the criterion variable is the gender projection index in the matched height judgment task (i.e., \(r(G,E)_{\text{matched}}\)). The variable \(r(G,E)_{\text{ecological}}\) is entered first to control for individual differences in prior belief concerning the magnitude of the height gap believed to exist between the two genders. Recall that someone who believes males to average 1 inch taller than females does not have to adjust their height estimates in the matched task to the same extent as someone who believes that the gap between the two genders in height is closer to 6 inches. Entering the \(r(G,E)_{\text{ecological}}\) variable first in the regression analysis is one other method of controlling for this differential magnitude of this gender-height difference in prior belief.
### Table 6

**Hierarchical Regression Analyses Predicting Projection of Prior Belief on the Matched Height Judgment Task**

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>$R^2$ Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$r(G,E)$, Ecological</td>
<td>.526</td>
<td>.277</td>
<td>43.98****</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Syllogisms, Belief Bias</td>
<td>.587</td>
<td>.068</td>
<td>11.82***</td>
<td>.306</td>
</tr>
<tr>
<td>2</td>
<td>AET</td>
<td>.614</td>
<td>.100</td>
<td>18.27****</td>
<td>-.372</td>
</tr>
<tr>
<td>2</td>
<td>Covariation Judgment†</td>
<td>.608</td>
<td>.086</td>
<td>15.42****</td>
<td>-.347</td>
</tr>
<tr>
<td>2</td>
<td>Confounded Variable††</td>
<td>.514</td>
<td>.003</td>
<td>0.51</td>
<td>.070</td>
</tr>
<tr>
<td>2</td>
<td>Cognitive Ability 1</td>
<td>.605</td>
<td>.089</td>
<td>16.08****</td>
<td>-.352</td>
</tr>
<tr>
<td>2</td>
<td>Cognitive Ability 2</td>
<td>.641</td>
<td>.134</td>
<td>25.98****</td>
<td>-.431</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001.

Note: N = 117, † = 116, †† = 107.

Listed next are alternative variables entered as a second step in the hierarchical regression. The question being addressed in this analysis is whether there are any variables that can account for unique variance in belief projection in the matched set once we partial out the variance accounted for by belief projection in the ecological set. As displayed in Table 6, once the magnitude of the gender stereotype is statistically controlled, significant additional variance in $r(G,E)_{matched}$ is explained by the syllogism belief bias index (belief bias score; 6.8% with a partial $r = .306$), the beta weight for argument quality in AET (see Table 6 for remaining values), and the beta weight for the $\Delta p$ statistic in the covariation task. That performance on these prior belief reasoning and judgment tasks have been repeatedly shown to be predictive of one another--despite the fact that they require distinctively different processing demands--suggests that we are looking at some true domain generality in the ability to reason about the world in the face of interfering prior belief. Table 6 also displays one other important fact...
Belief Bias
87

which has been repeatedly surfacing in all the previous analyses: Cognitive ability is a substantial predictor of the extent to which reasoning and judgment is either directly influenced by prior belief or performed well in the face of conflicting beliefs. In this case, Table 6 shows that both cognitive ability measures account for significant unique variance when entered in the second step of the hierarchical regression analysis.

**Does Cognitive Ability Alone Provide the Underlying Mechanism for the Observed Degree of Domain Generality?**

All previous analyses have consistently implicated cognitive ability as a significant factor in predicting or accounting for either a direct measure of belief bias, or a measure of reasoning well in the face of conflicting prior beliefs. The correlation between Cognitive Ability 1 and Cognitive Ability 2, as displayed in Table 5 with the syllogism belief bias score was -.445 and -.495 respectively (p < .001 for both), with the beta weight for argument quality in the AET the correlations were .449 and .392 (p < .001 for both), with the participants tracking of the Δp statistic in the covariation task the correlations were .493 and .505 respectively (p < .001 for both), and even with the belief congruence index for the confounded variable task—which with exception of the AET failed to correlate with all the other prior belief reasoning tasks—the cognitive ability measures nevertheless did correlate (r = -.224, p < .05; r = -.329 respectively). It is thus possible that the correlations and observed degree of domain generality between the prior belief tasks are ultimately subserved by general cognitive ability alone. This question is explored in the next series of hierarchical regression analyses.

Table 7 performs a similar hierarchical regression analysis to that performed in Table 6 except that this time around all the variance accounted for by both the cognitive ability measures are partialled out prior to entering any prior belief task index into the regression. Combined, both cognitive ability measures account for 13.8% of the variance in projecting the gender cue on the
matched height judgment task, i.e., \( r(G,E)_{\text{matched}} \), once the variance accounted for by \( r(G,E)_{\text{ecological}} \) is partialled out first. Note that by partialling the variance accounted for by the cognitive ability measures first, results in the belief bias score fail to account for any significant unique variance in \( r(G,E)_{\text{matched}} \). This need not, however, entail that belief bias in syllogistic reasoning is only linked to belief projection in the perceptual task via their apparent common dependency on cognitive ability. To demonstrate this point, the next variable entered in the hierarchical regression as an alternative fourth step is simply the total score correct on the 24 item syllogism test. This variable does predict additional \( r(G,E)_{\text{matched}} \) variance after first partialling out all the variance accounted for by cognitive ability. Although not a direct measure of belief bias (note the change of sign in the partial correlation), the syllogisms total score does capture the ability to evaluate correctly syllogistic arguments that contain conclusions which conflict with prior belief. Indeed we can see in the next step 4 alternatives that performance on the inconsistent syllogisms (which are included in the total syllogism score) account for a proportionally higher level of unique variance.

Recall that the syllogisms test consisted of three classes of syllogisms:

- Inconsistent are syllogisms in which the logical dimension is incongruent with the empirical dimension, syllogisms that have their empirical and logical dimensions congruent with one another are consistent, and syllogisms having no empirical dimension via the employment of nonsense words are neutral.

Entering the inconsistent syllogisms total score (a score which may range from 0--no items correct to 8--all items correct) as the fourth step results in not only accounting for a significant portion of additional variance, it also produces a slightly larger change in \( R^2 \), and a larger partial correlation than was the case when entering the entire syllogism test score. The inconsistent syllogism score gauges how well participants do at avoiding the intrusion of the
Table 7
Hierarchical Regression Analyses Predicting Projection of Prior Belief on the Matched Height Judgment Task with Cognitive Capacity Measures Entered First

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>$R^2$</th>
<th>Change</th>
<th>F Change</th>
<th>Partial $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>r(G,E), Ecological</td>
<td>.526</td>
<td>.277</td>
<td></td>
<td>43.98****</td>
<td>-</td>
</tr>
<tr>
<td>2.</td>
<td>Cognitive Ability 1</td>
<td>.605</td>
<td>.089</td>
<td></td>
<td>16.07****</td>
<td>-.352</td>
</tr>
<tr>
<td>3.</td>
<td>Cognitive Ability 2</td>
<td>.644</td>
<td>.049</td>
<td></td>
<td>9.44**</td>
<td>-.278</td>
</tr>
<tr>
<td>4.</td>
<td>Syllogisms, Belief Bias</td>
<td>.651</td>
<td>.009</td>
<td></td>
<td>1.67</td>
<td>.121</td>
</tr>
<tr>
<td>4.</td>
<td>Syllogisms, Total</td>
<td>.661</td>
<td>.022</td>
<td></td>
<td>4.29*</td>
<td>-.192</td>
</tr>
<tr>
<td>4.</td>
<td>Syllogisms, Incon.</td>
<td>.662</td>
<td>.023</td>
<td></td>
<td>4.61*</td>
<td>-.199</td>
</tr>
<tr>
<td>4.</td>
<td>Syllogisms, Con.</td>
<td>.650</td>
<td>.008</td>
<td></td>
<td>1.56</td>
<td>-.117</td>
</tr>
<tr>
<td>4.</td>
<td>Syllogisms, Neut.</td>
<td>.654</td>
<td>.013</td>
<td></td>
<td>2.47</td>
<td>-.147</td>
</tr>
<tr>
<td>4.</td>
<td>AET</td>
<td>.671</td>
<td>.035</td>
<td></td>
<td>7.16**</td>
<td>-.245</td>
</tr>
<tr>
<td>4.</td>
<td>Covariation Judgment††</td>
<td>.664</td>
<td>.014</td>
<td></td>
<td>2.82†</td>
<td>-.157</td>
</tr>
<tr>
<td>4.</td>
<td>Confounded Variable†††</td>
<td>.438</td>
<td>.004</td>
<td></td>
<td>0.57</td>
<td>-.072</td>
</tr>
</tbody>
</table>

† = p < .10, * = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001
Note: N = 117, †† = 116, ††† = 107; Incon. = Total score on inconsistent syllogisms, Con. = Total score on consistent syllogisms, Neut. = Total score on neutral syllogisms.

empirical dimension on evaluations of the logical dimension. Like the beta weight for argument quality in the AET, it is a measure of how well people reason in the face of conflicting belief. It does not, however, provide a means of capturing error in logic which is directly attributable to the influence of prior belief. This is the strength of the logic behind the belief bias score in which the inconsistent total is subtracted from the consistent total. Any errors arising in inconsistent syllogisms that are not attributable to prior belief influence should also be manifesting in the consistent syllogisms, and thus the subtraction of the inconsistent score from the consistent score leaves behind an index that should be
a function of only the directional asymmetry of prior belief influence. This strength of the belief bias score index, however, is accompanied by some serious reliability problems endemic to difference scores in general (Cronbach & Furby, 1970; Edwards, 1995, Thorndike & Hagen, 1977). This problematic nature of difference scores is further compounded by the fact that this particular derived difference score is calculated from only 16 items (8 consistent and 8 inconsistent). Table 7 furthermore shows that the syllogisms consistent variable accounted for no additional variance in the $r(G,E)_{matched}$, as was also the case with the syllogisms neutral score.

Although significantly reduced in absolute magnitude (cf., Table 6), the beta weight for argument quality in the AET continued to account for additional variance in the projection of the gender cue in the matched height judgment task viz., $r(G,E)_{matched}$, despite partialling out all the variance accounted for by both the cognitive ability measures. The beta weight for the the $\Delta p$ statistic on the covariation task was also substantially reduced as a predictor of unique variance when cognitive ability was entered into the hierarchical regression model first. Nevertheless, this index of judging bivariate relationships in the face of conflicting belief continued to account for additional variance in the projection of the gender cue in the matched height judgment task.

Table 8 and Table 9 display the results of two separate hierarchical regressions analyses performed on the respective criterion variables of belief bias score index and the syllogism total score index with the two cognitive ability measures entered into the model first. Only the correlations between actual target height ($T$) and height estimates ($E$) in the matched condition (i.e., $r(T,E)_{matched}$) and the SAD matched indices accounted for additional variance in the belief bias score index (see Table 8) once all the variance accounted for by cognitive ability is partialled out first. The prior belief tasks, however, were more
Successful at accounting for unique variance when the criterion variable was the syllogisms total score. As can be seen from Table 9, the beta weight for argument quality in the AET, the beta weight for the \( \Delta p \) statistic in the covariation task and the projection of the gender cue index in the matched height judgment task all account for additional variance in the syllogistic total score when entered as alternative third steps in a regression model with all the variance accounted for by cognitive ability already partialled out.

Table 8
Hierarchical Regression Analysis Predicting Belief Bias Score on Syllogism Test With Cognitive Capacity Measures Entered First

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>( R )</th>
<th>( R^2 ) Change</th>
<th>F Change</th>
<th>Partial ( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cognitive Ability 1</td>
<td>.445</td>
<td>.198</td>
<td>30.28****</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.513</td>
<td>.065</td>
<td>10.64**</td>
<td>- .284</td>
</tr>
<tr>
<td>3.</td>
<td>AET</td>
<td>.517</td>
<td>.004</td>
<td>0.68</td>
<td>-.075</td>
</tr>
<tr>
<td>3.</td>
<td>Covariation Judgment†</td>
<td>.520</td>
<td>.006</td>
<td>0.98</td>
<td>-.09</td>
</tr>
<tr>
<td>3.</td>
<td>Confounded Variable††</td>
<td>.508</td>
<td>.001</td>
<td>0.24</td>
<td>-.047</td>
</tr>
<tr>
<td>3.</td>
<td>( r(G,E) ) Matched†††</td>
<td>.504</td>
<td>.012</td>
<td>1.82</td>
<td>.126</td>
</tr>
<tr>
<td>3.</td>
<td>Gender Beta, Matched†††</td>
<td>.504</td>
<td>.012</td>
<td>1.87</td>
<td>.128</td>
</tr>
<tr>
<td>3.</td>
<td>Height Adjustment†††</td>
<td>.510</td>
<td>.018</td>
<td>2.77</td>
<td>-.155</td>
</tr>
<tr>
<td>3.</td>
<td>( r(T,E) ) Matched†††</td>
<td>.522</td>
<td>.030</td>
<td>4.74*</td>
<td>-.201</td>
</tr>
<tr>
<td>3.</td>
<td>SAD Matched†††</td>
<td>.548</td>
<td>.068</td>
<td>9.48**</td>
<td>.278</td>
</tr>
</tbody>
</table>

* = \( p < .05 \), ** = \( p < .01 \), *** = \( p < .001 \), **** = \( p < .0001 \)

Note: \( N = 124 \), \( \dagger = 123 \), \( \ddagger = 113 \), \( \dagger\ddagger = 117 \).
Table 9
Hierarchical Regression Analysis Predicting Total Syllogism Test Score With Cognitive Capacity Measures Entered First

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cognitive Ability 1</td>
<td>.573</td>
<td>.328</td>
<td>59.58****</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.608</td>
<td>.042</td>
<td>15.33**</td>
<td>.335</td>
</tr>
<tr>
<td>3.</td>
<td>AET</td>
<td>.658</td>
<td>.063</td>
<td>13.50***</td>
<td>.318</td>
</tr>
<tr>
<td></td>
<td>Covariation Judgment†</td>
<td>.662</td>
<td>.060</td>
<td>12.77***</td>
<td>.311</td>
</tr>
<tr>
<td>3.</td>
<td>Confounded Variable††</td>
<td>.621</td>
<td>.005</td>
<td>0.93</td>
<td>.092</td>
</tr>
<tr>
<td>3.</td>
<td>r(G,E) Matched††</td>
<td>.623</td>
<td>.026</td>
<td>4.83*</td>
<td>-.202</td>
</tr>
<tr>
<td>3.</td>
<td>Gender Beta, Matched†††</td>
<td>.623</td>
<td>.027</td>
<td>4.88*</td>
<td>-.203</td>
</tr>
<tr>
<td>3.</td>
<td>Height Adjustment†††</td>
<td>.604</td>
<td>.003</td>
<td>0.44</td>
<td>.062</td>
</tr>
<tr>
<td>3.</td>
<td>r(T,E) Matched†††</td>
<td>.614</td>
<td>.015</td>
<td>2.80</td>
<td>.155</td>
</tr>
<tr>
<td>3.</td>
<td>SAD Matched†††</td>
<td>.628</td>
<td>.033</td>
<td>6.10*</td>
<td>-.226</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001
Note: N = 124, † = 123, †† = 113, ††† = 117.

Table 10 summarizes the results of a series of hierarchical analyses exploring a more micro-level analysis of the three components of the syllogisms test total score. Identical hierarchical regressions analyses are performed alternating the inconsistent syllogisms score, consistent syllogisms score and neutral syllogisms score as the criterion variable. After partialling out all the variance in the inconsistent syllogism score which is accounted for by the cognitive ability measures (see middle column of Table 10), both the beta weight for argument quality in the AET and the beta weight for the Δp statistic in the covariation judgment accounted for unique variance when entered into the model as alternative third steps. All three of these indices (viz., syllogism inconsistent score, argument quality beta and Δp statistic beta) tap the ability to reason effectively independently of conflicting belief. The fact that they continue to account for unique variance even after all the variance attributable to cognitive
ability is partialled out, again suggests that a portion of the observed domain generality in reasoning independently of belief is not due only to a common reliance on cognitive ability.\footnote{Interestingly, both the beta weight for argument quality in the AET and the beta weight for the $\Delta p$ statistic in the covariation judgment also accounted for unique variance when entered as alternative third steps in a hierarchical regression analysis with either syllogism consistent score or syllogism neutral score as the criterion variable (see Table 10).}

Table 10
Hierarchical Regression Analyses Predicting Syllogism Problem-Type Performance with Cognitive Capacity Measures Forced in the Model First

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Consistent</th>
<th>Inconsistent</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cognitive Ability 1</td>
<td>.117****</td>
<td>.336****</td>
<td>.262****</td>
</tr>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.005</td>
<td>.068****</td>
<td>.021</td>
</tr>
<tr>
<td>3.</td>
<td>AET</td>
<td>.178**</td>
<td>.036**</td>
<td>.062***</td>
</tr>
<tr>
<td>3.</td>
<td>Covariation Judgment\dagger</td>
<td>.053**</td>
<td>.040**</td>
<td>.049**</td>
</tr>
<tr>
<td>3.</td>
<td>Confounded Variable\ddagger</td>
<td>.003</td>
<td>.006</td>
<td>.001</td>
</tr>
<tr>
<td>3.</td>
<td>r(G,E) Matched\ddagger\ddagger</td>
<td>.012</td>
<td>.026*</td>
<td>.020</td>
</tr>
<tr>
<td>3.</td>
<td>Gender Beta, Matched\ddagger\ddagger</td>
<td>.012</td>
<td>.026*</td>
<td>.020</td>
</tr>
<tr>
<td>3.</td>
<td>Height Adjustment\ddagger\ddagger</td>
<td>.001</td>
<td>.010</td>
<td>.001</td>
</tr>
<tr>
<td>3.</td>
<td>r(T,E) Matched\ddagger\ddagger</td>
<td>.001</td>
<td>.030*</td>
<td>.006</td>
</tr>
<tr>
<td>3.</td>
<td>SAD Matched\ddagger\ddagger</td>
<td>.000</td>
<td>.048**</td>
<td>.034*</td>
</tr>
</tbody>
</table>

R2 Change

* = p < .05, ** = p < .01, *** = p < .001, **** = p < .0001

Note: N = 124, † = 123, †† = 113, ††† = 117.

Finally, Table 11 presents a hierarchical regression analysis using the beta weight for argument quality in the AET as the criterion variable. Table 11 provides the additional information that the beta weight for the $\Delta p$ statistic in the covariation judgment task accounts for no additional variance in the AET index once the variability accounted for by cognitive ability is partialled out. The reverse is of course also true (AET does not predict additional variance on the covariation task above that already accounted for by cognitive ability) and this can be seen in
Table 12 which serves as a convenient layout for exploring a hierarchical regression analysis using the beta weight for the $\Delta p$ statistic as the criterion variable (the information in this table may be discerned from tables 8 through 11). It thus appears that the correlation between the ability to reason well in the face of conflicting prior beliefs on the AET and the covariation judgment tasks may have in fact been a product of their common reliance on general cognitive ability.

Table 11
Hierarchical Regression Analyses Predicting Beta Weight for Argument Quality on the Argument Evaluation Test with Cognitive Capacity Measures Entered First

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>$R$</th>
<th>$R^2$ Change</th>
<th>$F$ Change</th>
<th>Partial $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.461</td>
<td>.011</td>
<td>1.72</td>
<td>.119</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Belief Bias</td>
<td>.466</td>
<td>.004</td>
<td>0.68</td>
<td>-0.75</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Total</td>
<td>.541</td>
<td>.079</td>
<td>13.50***</td>
<td>.318</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Incon. Total</td>
<td>.510</td>
<td>.047</td>
<td>7.72**</td>
<td>.246</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Con. Total</td>
<td>.512</td>
<td>.050</td>
<td>8.10**</td>
<td>.251</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Neut. Total</td>
<td>.530</td>
<td>.068</td>
<td>11.31***</td>
<td>.293</td>
</tr>
<tr>
<td>3.</td>
<td>Covariation Judgment†</td>
<td>.460</td>
<td>.000</td>
<td>0.002</td>
<td>.004</td>
</tr>
<tr>
<td>3.</td>
<td>Confounded Variable‡‡</td>
<td>.438</td>
<td>.004</td>
<td>.572</td>
<td>-.072</td>
</tr>
<tr>
<td>3.</td>
<td>$r$(G,E) Matched‡‡‡</td>
<td>.462</td>
<td>.034</td>
<td>4.91*</td>
<td>-.204</td>
</tr>
<tr>
<td>3.</td>
<td>Gender Beta, Matched‡‡‡</td>
<td>.462</td>
<td>.035</td>
<td>4.97*</td>
<td>-.205</td>
</tr>
<tr>
<td>3.</td>
<td>Height Adjustment‡‡‡</td>
<td>.440</td>
<td>.014</td>
<td>3.29</td>
<td>.168</td>
</tr>
</tbody>
</table>

* = $p < .05$, ** = $p < .01$, *** = $p < .001$, **** = $p < .0001$
Note: $N = 124$, †= 123, ‡‡ = 113, ‡‡‡ = 117
Table 12

Hierarchical Regression Analyses Predicting Beta Weight for the Delta p Statistic in the Covariation Judgment task with Cognitive Capacity Measures Entered First

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cognitive Ability 1</td>
<td>.493</td>
<td>.243</td>
<td>38.81***</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.539</td>
<td>.047</td>
<td>8.03**</td>
<td>.251</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Belief Bias</td>
<td>.544</td>
<td>.006</td>
<td>0.98</td>
<td>-.09</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Total</td>
<td>.599</td>
<td>.069</td>
<td>12.77**</td>
<td>.311</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Incon. Total</td>
<td>.581</td>
<td>.048</td>
<td>8.56**</td>
<td>.259</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Con. Total</td>
<td>.578</td>
<td>.044</td>
<td>7.72**</td>
<td>.247</td>
</tr>
<tr>
<td>3.</td>
<td>Syllogisms, Neut. Total</td>
<td>.582</td>
<td>.049</td>
<td>8.79**</td>
<td>.262</td>
</tr>
<tr>
<td>3.</td>
<td>AET</td>
<td>.546</td>
<td>.008</td>
<td>1.34</td>
<td>.105</td>
</tr>
<tr>
<td>3.</td>
<td>Confounded Variable†</td>
<td>.512</td>
<td>.001</td>
<td>0.14</td>
<td>.036</td>
</tr>
<tr>
<td>3.</td>
<td>r(G,E) Matched††</td>
<td>.560</td>
<td>.009</td>
<td>1.41</td>
<td>-.112</td>
</tr>
<tr>
<td>3.</td>
<td>Gender Beta, Matched††</td>
<td>.560</td>
<td>.009</td>
<td>1.40</td>
<td>-.111</td>
</tr>
<tr>
<td>3.</td>
<td>Height Adjustment††</td>
<td>.560</td>
<td>.009</td>
<td>1.41</td>
<td>.111</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .025, *** = p < .01, **** = p < .001
Note: N = 123, † = 113, †† = 116

The Role of Thinking Dispositions on Prior Belief Task Performance

Table 13 displays the intercorrelations of all the thinking disposition scales. With very little exception (none of the exceptions approaching significance), the direction of the correlations between these diverse scales ran in the anticipated directions. Thus, for example, the flexible thinking scale correlated positively with Openness-Ideas, and Openness-Values whereas it correlated negatively with Absolutism, Dogmatism, Need for Closure, Belief Identification, Categorical Thinking, Paranormal Beliefs etc. Because thinking dispositions lending
Table 13

Intercorrelations Among the Primary Thinking Disposition Variables

(N = 124)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Flexible Thinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Openness-Ideas (NEO)</td>
<td>.250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Openness-Values (NEO)</td>
<td>.459</td>
<td>.171</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Absolutism</td>
<td>-.424</td>
<td>-.347</td>
<td>-.623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Dogmatism</td>
<td>-.320</td>
<td>-.101</td>
<td>-.490</td>
<td>.513</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Need for Closure</td>
<td>-.193</td>
<td>-.149</td>
<td>-.147</td>
<td>.356</td>
<td>.200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Belief Identification</td>
<td>-.527</td>
<td>-.245</td>
<td>-.523</td>
<td>.549</td>
<td>.378</td>
<td>.148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Categorical Thinking</td>
<td>-.254</td>
<td>-.060</td>
<td>-.456</td>
<td>.557</td>
<td>.550</td>
<td>.167</td>
<td>.403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Paranormal Beliefs</td>
<td>-.218</td>
<td>.039</td>
<td>-.003</td>
<td>.205</td>
<td>.008</td>
<td>-.091</td>
<td>.245</td>
<td>.119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Concept of Luck</td>
<td>-.235</td>
<td>-.166</td>
<td>.067</td>
<td>.098</td>
<td>-.064</td>
<td>-.037</td>
<td>.271</td>
<td>.044</td>
<td>.629</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Deliberation (NEO)</td>
<td>.078</td>
<td>.215</td>
<td>-.021</td>
<td>.096</td>
<td>.039</td>
<td>.413</td>
<td>-.087</td>
<td>.023</td>
<td>-.046</td>
<td>-.137</td>
<td>-.220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. CFC</td>
<td>.291</td>
<td>.484</td>
<td>.079</td>
<td>-.190</td>
<td>-.039</td>
<td>.055</td>
<td>-.067</td>
<td>-.027</td>
<td>-.178</td>
<td>-.200</td>
<td>-.105</td>
<td>.257</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Impulsiveness (NEO)</td>
<td>.076</td>
<td>-.142</td>
<td>.103</td>
<td>-.063</td>
<td>-.053</td>
<td>-.238</td>
<td>-.081</td>
<td>.007</td>
<td>.156</td>
<td>.126</td>
<td>.171</td>
<td>-.440</td>
<td>-.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Counterfactual Thinking</td>
<td>.272</td>
<td>.038</td>
<td>.206</td>
<td>-.252</td>
<td>-.071</td>
<td>-.044</td>
<td>-.293</td>
<td>-.200</td>
<td>-.197</td>
<td>-.181</td>
<td>-.018</td>
<td>-.051</td>
<td>.171</td>
<td>.168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Faith in Intuition</td>
<td>-.085</td>
<td>.009</td>
<td>.022</td>
<td>.196</td>
<td>.327</td>
<td>.093</td>
<td>.132</td>
<td>.088</td>
<td>.206</td>
<td>.181</td>
<td>.156</td>
<td>-.075</td>
<td>-.121</td>
<td>.126</td>
<td>.029</td>
<td></td>
</tr>
<tr>
<td>17. Head Over Heart</td>
<td>.082</td>
<td>.320</td>
<td>.053</td>
<td>-.120</td>
<td>-.012</td>
<td>.259</td>
<td>-.194</td>
<td>.000</td>
<td>-.072</td>
<td>-.228</td>
<td>-.113</td>
<td>.510</td>
<td>.321</td>
<td>-.349</td>
<td>-.016</td>
<td>-.113</td>
</tr>
</tbody>
</table>

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)

Note: CFC = Consideration of Future Consequences scale
themselves to actively open-minded thinking are of particular interest to this research, a larger composite scale of actively open-minded thinking was derived from the associated scales. The construction of the actively open-minded thinking scale—along with three other thinking disposition scale composites to be described—was largely informed through the results of a principal components analysis conducted on the 17 subscales as displayed in Table 13. Several orthogonal and oblique rotations converged on a specified three component model—obtained after a varimax rotation—as the most conceptually interpretable and parsimonious summary of the covariation data. All three components had eigenvalues > 1. The results of this analysis is summarized in Table 14. The first component displayed on Table 14 (which accounted for 25.4% of the variance) received high loadings from Flexible Thinking, Openness-Ideas, Openness-values, Absolutism, Dogmatism, Need for Closure, Belief Identification, and Categorical Thinking. This factor is interpreted here to be that dimension tapping the thinking disposition construct of actively open-minded thinking which is of central importance to this investigation. At one end of this important dimension is cognitive flexibility and openness to belief change versus cognitive rigidity or resistance to belief change found on the other end of this dimension. The second factor (which accounted for 18.1% of the variance) is interpreted here to be tapping conscientiousness. This interpretation seems justified with consideration being given to the nature of the scales loading on this second factor (e.g., Deliberation, Consideration of Future Consequences). The Openness-Ideas scale loaded on the second factor higher than the first factor (.558 vs .354), but because the Openness-Ideas scale is a measure of a dimension argued here to be of importance to actively open-minded thinking, this aspect of the principal components analysis was set aside on conceptual grounds. The Openness-Ideas scale was, instead, included in the construction of the actively open-minded
thinking composite.

Table 14
Component Loadings for Thinking Disposition Variables after Varimax Rotation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AOT</td>
<td>-.626</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Openness-Ideas</td>
<td>-.354</td>
<td>.558</td>
<td></td>
</tr>
<tr>
<td>3. Openness-Values</td>
<td>-.789</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Absolutism</td>
<td>.841</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Dogmatism</td>
<td>.700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Need for Closure</td>
<td>.429</td>
<td>.401</td>
<td></td>
</tr>
<tr>
<td>7. Belief Identification</td>
<td>.701</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Categorical Thinking</td>
<td>.669</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Paranormal Thinking</td>
<td></td>
<td></td>
<td>.868</td>
</tr>
<tr>
<td>10. Concept of Luck</td>
<td></td>
<td></td>
<td>.802</td>
</tr>
<tr>
<td>11. Superstitious Thinking</td>
<td></td>
<td></td>
<td>.728</td>
</tr>
<tr>
<td>12. Deliberation</td>
<td></td>
<td>.773</td>
<td></td>
</tr>
<tr>
<td>13. CFC</td>
<td></td>
<td>.579</td>
<td></td>
</tr>
<tr>
<td>14. Impulsiveness</td>
<td></td>
<td>-.617</td>
<td></td>
</tr>
<tr>
<td>15. Counter Factual Thinking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Faith in Intuition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Head over Heart</td>
<td></td>
<td>.768</td>
<td></td>
</tr>
</tbody>
</table>

% variance accounted for
25.4% 18.1% 11.6%

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

The actively open-minded thinking (AOT) composite scale was formed by first standardizing the total scores on the Flexible Thinking Scale, Openness Ideas Scale, Openness-Values Scale, Absolutism Scale, Dogmatism Scale, Need for Closure Scale, Belief Identification Scale, Categorical Thinking Scale (viz., all scales loading higher than .350 on factor 1) and then summing these scores into a composite scale of actively open-minded thinking. Although the AOT composite
scale is of central importance to this research, two other scales were also derived on the basis of the principal components analysis summarized on Table 14. The scales having a loading greater than .350 on factor 2 (with the exception of Openness-Ideas and Need for Closure) were subjected to the same procedure above to form a composite scale for conscientiousness. Similarly, the scales that loaded on factor 3 were used to derive a composite scale of superstitious thinking (again labeled as such with consideration of the nature of the scales loading on this third factor). The superstitious thinking factor accounted for a total of 11.6% of the variance. The three generated factors thus accounted for a combined 55.1% of the variance.

Table 15 displays the zero-order correlations between the three thinking disposition composite scales and various prior belief task performance indices. Several of the prior belief task indices correlated mildly to moderately with one or more of the thinking disposition composite scales. Importantly, this was particularly true with the AOT composite scale. With the exception of the belief congruency index for the confounded variable task and the indices for the height judgment tasks, the AOT composite scale correlated with all the remaining prior belief task indices. Thus task performance on the prior belief tasks was by-in-large correlated in the anticipated direction with actively open-minded thinking (AOT). Although not as successful as the AOT, the superstitious thinking scale (ST) also correlated mildly with some of the syllogism test indices, the covariation judgment task index and moderately with the argument evaluation test (AET). Least successful was the conscientiousness composite which only reached significant correlation levels with the AET and the total consistent syllogism score.

Table 15 also displays the correlation between the thinking disposition composite scales and the two measures of cognitive ability. The AOT was
correlated with both cognitive ability measures, and this was also true to a lesser degree with the superstitious thinking composite. The conscientiousness composite was not found to correlate with either cognitive ability measure.

Table 15
Zero-Order Correlations Between Various Indices on the Prior Belief Tasks with the Composite Thinking Disposition Scales for Actively Open-Minded Thinking (AOT), Conscientiousness (Cons), Superstitious Thinking (ST) and the Impression Management Scale

<table>
<thead>
<tr>
<th>Variable</th>
<th>Composite Scale</th>
<th>Impression Management</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AOT</td>
<td>Cons</td>
</tr>
<tr>
<td>Syllogisms, Belief Bias</td>
<td>-.315***</td>
<td>.072</td>
</tr>
<tr>
<td>Syllogisms, Total</td>
<td>.408***</td>
<td>.141</td>
</tr>
<tr>
<td>Syllogisms, Incon. Total</td>
<td>.422***</td>
<td>.076</td>
</tr>
<tr>
<td>Syllogisms, Con. Total</td>
<td>.263**</td>
<td>.244**</td>
</tr>
<tr>
<td>Syllogisms, Neut. Total</td>
<td>.339***</td>
<td>.095</td>
</tr>
<tr>
<td>AFT</td>
<td>.436***</td>
<td>.182*</td>
</tr>
<tr>
<td>Covariation Judgment†</td>
<td>.241**</td>
<td>.050</td>
</tr>
<tr>
<td>Confounded Variable†††</td>
<td>-.080</td>
<td>.133</td>
</tr>
<tr>
<td>r(G,E) Ecological††††</td>
<td>.057</td>
<td>-.136</td>
</tr>
<tr>
<td>r(G,E) Matched††††††</td>
<td>-.130</td>
<td>-.119</td>
</tr>
<tr>
<td>Height Adjustment†††††</td>
<td>.119</td>
<td>-.078</td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td>.358***</td>
<td>.104</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td>.323***</td>
<td>-.010</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001
Note: N = 124, † = 123, †† = 113, ††† = 117

Is the Association Between Participants Giving the Normative Response on the Prior Belief Tasks and the Thinking Disposition Measures a Product of Those Participants Giving Socially Desirable Responses?

An alternative explanation of the pattern of correlations between the prior belief tasks and thinking disposition composite scales is that participants
performing at the top end of each of the tasks were also more likely to give socially desirable responses than those performing at lower levels. If this were to be the case then the correlation between the tasks and the thinking disposition scales is merely a reflection of this selective response bias, viz., endorsing items seen as socially desirable and vice versa. This issue is also addressed in Table 15. The fourth column of Table 15 displays the correlations between the same prior belief task indices and the total score obtained in the administered impression management scale—a scale included in the battery for the purpose of detecting possible social desirability response biases. Two indices of the prior belief tasks were found to correlate to a significant degree with scores on the Impression Management Scale: The beta weight for argument quality on the AET, and \( r(G,E)_{\text{ecological}} \). The Impression Management score was found to correlate negatively with the performance index on the AET. Thus rather than giving more socially desirable responses, those participants reasoning well in the face of conflicting belief on the AET were not only more likely to score in the anticipated direction on the three composite thinking disposition scales, they were also less likely to constrain their responses to that which may seem as socially desirable.

In fact, a quick inspection of column four of Table 15 suggests that this was by-and-large the general trend. Although not reaching levels of statistical significance, it seems that it was participants who did less well on the prior belief tasks that were more prone to score higher on the Impression Management Scale. This was also true with the only other index to reach significance, i.e., \( r(G,E)_{\text{ecological}} \). Recall that participants who projected the gender cue in the ecological height judgment task were more accurate in their height judgments, i.e., \( r(G,E)_{\text{ecological}} \) was positively correlated with the accuracy measures used. Projection of gender in the ecological height judgment task—although not significant with any of the composite thinking disposition scales—was negatively
correlated with impression management. To summarize, it does not seem that the positive correlations between prior belief task indices and the composite thinking disposition scales was merely a by-product of the high-achievers giving more socially desirable responses. If anything, there lies some small suggestion that it was the participants who performed less competently on the prior belief tasks that were more likely to constrain their responses to socially desirable ones.

**Do Thinking Dispositions Account for any Additional Variance in the Prior Belief Tasks Once the Variance Accounted for by Cognitive Ability is Partialled Out?**

Table 15 demonstrated that the composite thinking disposition scales are correlated with some of the prior belief reasoning tasks. Of particular importance to the present research is the nature of this relationship between the AOT composite scale and the prior belief reasoning tasks. In order to explore whether the AOT composite accounts for any additional variance after all the variance accounted for by cognitive ability is first partialed out, a series of hierarchical regression analyses were performed on the prior belief indices exhibiting correlations with the AOT composite. The results varied greatly as a function of which prior belief task is under consideration. Table 16 displays the results of a hierarchical analysis using the criterion variable found to correlate most heavily with the AOT, viz., the beta weight for argument quality on the AET. Combined, the two cognitive ability measures accounted for 21.3% of the variance in this AET index. Entered next as alternative 3rd steps in the hierarchical regression is the AOT composite, the conscientiousness composite, and the superstitious thinking composite. All three composite scales accounted for additional variance on the beta weight for argument quality after all the variance accounted for by cognitive ability is partialled out. The AOT composite was of particular note here accounting for an additional 8.1% of the variance.
Table 16
Hierarchical Regression Analyses Predicting Beta Weight for Argument Quality on the AET with Cognitive Capacity measures Entered First followed by Thinking Disposition Scales

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R^2 Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.461</td>
<td>.011</td>
<td>1.72</td>
<td>.119</td>
</tr>
<tr>
<td>3.</td>
<td>AOT</td>
<td>.542</td>
<td>.081</td>
<td>13.81***</td>
<td>.321</td>
</tr>
<tr>
<td>3.</td>
<td>Con</td>
<td>.485</td>
<td>.022</td>
<td>3.52†</td>
<td>.169</td>
</tr>
<tr>
<td>3.</td>
<td>ST</td>
<td>.504</td>
<td>.041</td>
<td>6.63*</td>
<td>-.229</td>
</tr>
</tbody>
</table>

† = p < .10, * = p < .05, ** = p < .01, *** = p < .001
Note: N = 124; AOT = actively open-minded thinking composite, Con = Conscientiousness composite, ST = superstitious thinking composite.

Not shown in Table 16 is the further finding that even in the case where the AOT composite scale is forced into the hierarchical regression as a third step, the superstitious thinking composite scale would account for additional variance as an alternative fourth step (partial r = -.229, p < .05). Although not statistically significant at the p = .05 level, the conscientiousness composite scale entered as an alternative fourth step also displayed a similar trend (partial r = .162, p < .10). Both the superstitious thinking composite scale and the conscientiousness composite scale explained no additional variance on the other prior belief task indices and are thus not shown for the remaining analyses.

The next series of hierarchical analyses are summarized in Table 17. These hierarchical analyses explore the contribution of the AOT in accounting for unique variance on various indices on the syllogisms test. The first analysis shows that after partialling out all the variance accounted for by cognitive capacity, the AOT composite scale accounted for some additional variance in the syllogism belief bias score—although this missed statistical significance at the p < .05 level (p = .07). As shown in the next hierarchical analysis, the AOT composite
Table 17

Hierarchical Regression Analyses Predicting Various Performance Indices on the Syllogistic Reasoning Test with Cognitive Capacity Measures Entered First Followed by Thinking Disposition Scales

<table>
<thead>
<tr>
<th>Syllogisms, Belief Bias</th>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Cognitive Ability 1</td>
<td>.445</td>
<td>.198</td>
<td>30.21****</td>
<td>-.284</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Cognitive Ability 2</td>
<td>.513</td>
<td>.065</td>
<td>10.64****</td>
<td>-.248</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AOT</td>
<td>.532</td>
<td>.020</td>
<td>3.31†</td>
<td>-.164</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syllogisms, Total</th>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Cognitive Ability 1</td>
<td>.573</td>
<td>.328</td>
<td>59.58****</td>
<td>.248</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Cognitive Ability 2</td>
<td>.608</td>
<td>.042</td>
<td>7.96**</td>
<td>.249</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AOT</td>
<td>.639</td>
<td>.039</td>
<td>7.91**</td>
<td>.264</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syllogisms, Inconsistent Total</th>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Cognitive Ability 1</td>
<td>.579</td>
<td>.336</td>
<td>61.66****</td>
<td>.321</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Cognitive Ability 2</td>
<td>.636</td>
<td>.068</td>
<td>13.93****</td>
<td>.321</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AOT</td>
<td>.668</td>
<td>.042</td>
<td>8.97**</td>
<td>.264</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syllogisms, Consistent Total</th>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Cognitive Ability 1</td>
<td>.342</td>
<td>.117</td>
<td>16.11****</td>
<td>.161</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Cognitive Ability 2</td>
<td>.349</td>
<td>.005</td>
<td>3.20†</td>
<td>.154</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AOT</td>
<td>.378</td>
<td>.021</td>
<td>2.90†</td>
<td>.154</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syllogisms, Neutral Total</th>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Cognitive Ability 1</td>
<td>.512</td>
<td>.262</td>
<td>43.40****</td>
<td>.169</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Cognitive Ability 2</td>
<td>.532</td>
<td>.021</td>
<td>3.55†</td>
<td>.169</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AOT</td>
<td>.554</td>
<td>.024</td>
<td>4.03*</td>
<td>.180</td>
</tr>
</tbody>
</table>

† = p < .10, * = p < .05, ** = p < .01, *** = p < .001
Note: N = 124

scale was more substantial when accounting for additional variance on the
syllogism test total score. Once again a consideration of the components of the syllogism total score (inconsistent, consistent and neutral syllogisms) demonstrate that it was the performance on the inconsistent syllogisms that were particularly telling. The next three criterion variables in the series of hierarchical regressions shown in Table 17 are the inconsistent, consistent and neutral syllogisms scores. After partialling out all the variance accounted for by cognitive ability, the AOT composite produced a slightly larger $R^2$ change in the inconsistent syllogism score than the syllogism total score (.042 vs .039)--even though the total score index has more items. Furthermore, the AOT did not account for as large a proportion of unique variance in the consistent and neutral syllogism scores. Although not accounting for additional variance on the consistent syllogism score, the AOT did account for some unique variance on the syllogism neutral score. Thus on both the AET and the syllogisms test, the AOT was found to be a good predictor of reasoning--above and beyond what cognitive ability could explain--when that reasoning was performed in the presence of conflicting prior beliefs.

The next three hierarchical analyses displayed in Table 18 show that once all the possible variance accounted for by cognitive ability was statistically partialled out, the AOT did not account for any additional variance in the covariation task, confounded variable task, the height adjustment score index or the matched height judgment task.

Thus the extent to which the AOT accounted for additional variance above and beyond what can be already attributed to cognitive ability is very much dependent on the type of task under consideration. The AOT explained some additional variance on various indices of syllogistic reasoning. But it was the argument evaluation test in which the AOT proved to be a powerful explanatory construct.
Table 18
Hierarchical Regression Analyses Predicting Performance Indices on the Covariation Judgment Task, Confounded Variable Task and Height Judgment Task with Cognitive Capacity Measures Entered First Followed by Thinking Disposition Scales

<table>
<thead>
<tr>
<th>Covariation Judgment (N = 123)</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cognitive Ability 1</td>
<td>.493</td>
<td>.243</td>
<td>38.81****</td>
<td>38.81****</td>
<td>.251</td>
</tr>
<tr>
<td>2. Cognitive Ability 2</td>
<td>.539</td>
<td>.047</td>
<td>8.03**</td>
<td>8.03**</td>
<td>.251</td>
</tr>
<tr>
<td>3. AOT</td>
<td>.541</td>
<td>.002</td>
<td>0.35</td>
<td>0.35</td>
<td>.054</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confounded Variable (N = 113)</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cognitive Ability 1</td>
<td>.224</td>
<td>.050</td>
<td>5.87*</td>
<td>5.87*</td>
<td>-.247</td>
</tr>
<tr>
<td>2. Cognitive Ability 2</td>
<td>.329</td>
<td>.058</td>
<td>7.16**</td>
<td>7.16**</td>
<td>.251</td>
</tr>
<tr>
<td>3. AOT</td>
<td>.330</td>
<td>.001</td>
<td>0.47</td>
<td>0.47</td>
<td>.021</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Height Adjustment Score (N = 117)</th>
<th>Variable</th>
<th>R</th>
<th>R² Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cognitive Ability 1</td>
<td>.347</td>
<td>.121</td>
<td>15.76****</td>
<td>15.76****</td>
<td>.325</td>
</tr>
<tr>
<td>2. Cognitive Ability 2</td>
<td>.462</td>
<td>.091</td>
<td>13.48***</td>
<td>13.48***</td>
<td>.325</td>
</tr>
<tr>
<td>3. AOT</td>
<td>.463</td>
<td>.002</td>
<td>0.08</td>
<td>0.08</td>
<td>-.026</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001
Note: AOT = actively open-minded thinking composite.

A further converging examination of these covariance relationships among performance on the AET, cognitive ability measures, and actively open-minded thinking dispositions can be seen in the summary of a principal components analysis displayed in Table 19. The resulting component loadings for all the variables using a varimax rotation in a specified 2-factor model is displayed in Table 19. The first component, which accounted for 38.4% of the variance, received high loadings from six of the thinking disposition scales whereas the
second factor received high loadings from both the cognitive ability measures and one thinking disposition scale. For this reason the first factor is interpreted here to be tapping dispositions toward actively open-minded thinking and the second factor to be tapping general cognitive ability. Of particular note here is that the beta weight for argument quality on the AET loads highly on both the cognitive ability and thinking dispositional factors. Thus once again the data suggests that performance on the AET is a function of two distinct levels of analysis in cognitive theory.

Table 19
Component Loadings on a Two Factor Model for all Actively Open-minded Thinking Disposition Subscales, Cognitive Ability Measures and the AET Beta for Argument Quality After Varimax Rotation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AET Beta for Argument Quality</td>
<td>-.410</td>
<td>.503</td>
</tr>
<tr>
<td>GCA 1</td>
<td>-----</td>
<td>.873</td>
</tr>
<tr>
<td>GCA 2</td>
<td>-----</td>
<td>.871</td>
</tr>
<tr>
<td>Flexible Thinking</td>
<td>-.541</td>
<td>.369</td>
</tr>
<tr>
<td>Openness-Ideas</td>
<td>-----</td>
<td>.509</td>
</tr>
<tr>
<td>Openness-Values</td>
<td>-.782</td>
<td>-----</td>
</tr>
<tr>
<td>Absolutism</td>
<td>.793</td>
<td>-----</td>
</tr>
<tr>
<td>Dogmatism</td>
<td>.746</td>
<td>-----</td>
</tr>
<tr>
<td>Categorical Thinking</td>
<td>.765</td>
<td>-----</td>
</tr>
<tr>
<td>Belief Identification</td>
<td>.675</td>
<td>-----</td>
</tr>
<tr>
<td>Need for Closure</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>% variance accounted for</td>
<td>38.4%</td>
<td>14.5%</td>
</tr>
</tbody>
</table>

Component loadings lower than .350 have been eliminated

Note: Component 1 and component 2 interpreted as actively open-minded thinking dispositions and cognitive ability respectively.
How Much Shared Variance Exists Between Thinking Dispositions and Cognitive Ability Measures?

The hierarchical analyses displayed in Tables 16-18 consistently forced the cognitive ability measures into the hierarchical regression prior to entering the AOT composite. From the standpoint of the utility of the AOT composite scale, the results thus represent a worst case scenario. All of the common variance between the AOT and the cognitive ability measures is attributed to the latter. Although in theory this is a possibility, it is more likely that some of this common variance can be attributed to the AOT scale. Table 20 puts this issue into some perspective through the use of a series of commonality analyses.

The criterion variable in the first commonality analysis is the beta weight for argument quality in the AET. The total amount of variance explained by both cognitive ability measures and the three thinking disposition composite scales (i.e., the AOT, ST, and Cons.) was 33.8%. A total of 14% of the accounted variance was common to the thinking disposition scale and the two cognitive ability measures. Recall that in the hierarchical analysis displayed in Table 16, all of this shared variance was attributed to the cognitive ability measures alone--a worst case scenario. In terms of accounting for unique variance on the AET, the combination of the three thinking disposition composite scales was more substantial than both cognitive ability measures combined. While the two cognitive ability measures accounted for a total of 7.3% unique variance in the beta weight for argument quality on the AET, the three thinking disposition composite scales accounted for 12.5% of unique variance. Since the superstitious thinking composite and the conscientiousness composite scales accounted for little variance on the other variables, only the AOT was used in the remaining commonality analyses displayed in Table 20.

The next commonality analysis displayed on Table 20 uses the belief bias score as the criterion variable. It is clear this time around that the cognitive
Table 20
Commonality Analyses of Thinking Disposition Measures and Cognitive Ability Measures for all Prior Belief Tasks

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Unique</th>
<th>Common</th>
<th>Total Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thinking Dispositions</strong></td>
<td>.125</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.073</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>.338</strong></td>
</tr>
<tr>
<td><strong>Syllogism, Belief Bias</strong></td>
<td>Unique</td>
<td>Common</td>
<td>Total Variance Explained</td>
</tr>
<tr>
<td>AOT</td>
<td>.020</td>
<td>.079</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.184</td>
<td>.079</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>.283</strong></td>
</tr>
<tr>
<td><strong>Syllogism, Total Score</strong></td>
<td>Unique</td>
<td>Common</td>
<td>Total Variance Explained</td>
</tr>
<tr>
<td>AOT</td>
<td>.039</td>
<td>.128</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.242</td>
<td>.128</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>.409</strong></td>
</tr>
<tr>
<td><strong>Syllogism, Inconsistent Total</strong></td>
<td>Unique</td>
<td>Common</td>
<td>Total Variance Explained</td>
</tr>
<tr>
<td>AOT</td>
<td>.042</td>
<td>.136</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.268</td>
<td>.136</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>.446</strong></td>
</tr>
<tr>
<td><strong>Syllogism, Consistent Total</strong></td>
<td>Unique</td>
<td>Common</td>
<td>Total Variance Explained</td>
</tr>
<tr>
<td>AOT</td>
<td>.021</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.074</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>.143</strong></td>
</tr>
<tr>
<td><strong>Syllogism, Neutral Total</strong></td>
<td>Unique</td>
<td>Common</td>
<td>Total Variance Explained</td>
</tr>
<tr>
<td>AOT</td>
<td>.024</td>
<td>.091</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.192</td>
<td>.091</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>.307</strong></td>
</tr>
</tbody>
</table>

Note: Table continued on next page. Thinking Dispositions = Combined variance accounted for by the actively-open-minded thinking composite, conscientiousness composite, and superstitious thinking composite; Cognitive Ability = Combined variance accounted for by both Cognitive Ability 1 and Cognitive Ability 2 composite scales.
Table 20 (continued)

**Dependent Variable:**

**Beta Weight for Delta P Statistic on the Covariation Judgment Task**

<table>
<thead>
<tr>
<th></th>
<th>Unique</th>
<th>Common</th>
<th>Total Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOT</td>
<td>.002</td>
<td>.056</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.234</td>
<td>.056</td>
<td>.292</td>
</tr>
</tbody>
</table>

**Height Adjustment Index for Matched Height Judgment Task**

<table>
<thead>
<tr>
<th></th>
<th>Unique</th>
<th>Common</th>
<th>Total Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOT</td>
<td>.000</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.200</td>
<td>.014</td>
<td>.214</td>
</tr>
</tbody>
</table>

**Belief Congruence Score on Confounded Variable Task**

<table>
<thead>
<tr>
<th></th>
<th>Unique</th>
<th>Common</th>
<th>Total Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOT</td>
<td>.001</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.103</td>
<td>.005</td>
<td>.109</td>
</tr>
</tbody>
</table>

Note: Cognitive Ability = variance accounted for by both Cognitive Ability 1 and Cognitive Ability 2 composite scales.

ability measures account for an appreciably larger portion of unique variance.

The common variance accounted for by both the AOT and the cognitive ability measures was 7.9% on the belief bias score. Again, this common variance was previously attributed solely to cognitive ability in the analyses displayed in Table 17. A total of 12.8% of the accounted for variance on the syllogism test total score was common to the AOT and cognitive ability. A closer examination of the components of this syllogism total score revealed that while AOT accounted for 4.2% of unique variance, an additional 13.6% of the variance on the Inconsistent total was common to the AOT and cognitive ability. Once again cognitive ability played a less dominant role in accounting for both unique and common variance on the other two components of the syllogisms test total score (consistent and neutral total scores).

Using the three indices for the remaining tasks as the criterion variables
(the next three commonality analyses on Table 20), the AOT accounted for virtually no unique variance and shared very little (5.6% for the covariation judgment task) to essentially no common variance with the cognitive ability measures cf., .014 and .005 for the height adjustment score index on the matched height judgment task, and belief congruence score index on the confounded variable task respectively.

**Further Explorations of Thinking Dispositions: Epistemic Development in Higher Education**

Another interesting issue which is possible to address with the present data is the role higher education may or may not play in the development of thinking dispositions. In outlining his model of epistemic development, William Perry (1970), for example, noted that university students enrolled in the humanities often complained that the “authorities did not seem to be clear about either the right answers or about how one was to go about one’s work in obtaining them”. Perry believed that personal epistemology developed in a fixed progression of stages so that even those above-mentioned students holding beliefs in absolute knowledge may progress to higher and more developmentally advanced stages of personal epistemology. Baron (1991) makes the tie to educational experience explicit by suggesting that thinking dispositions are “learned tendencies to behave in certain ways” (p. 118, my italics). Schommer (1990) also argues that epistemic beliefs are influenced by education (and the home) and has also suggested that they influence academic performance (Schommer, 1993; see also Schoenfeld, 1983, 1985; Dweck & Leggett, 1988).

Thus a basic question is simply whether the present data shows any differences in thinking dispositions as a function of year enrolled in university. Participants were first grouped into one of four groups: year 1 students (N = 45), year 2 students (N = 26), year 3 students (N = 20), and year 4 or more students (N = 31). A series of one way ANOVAs was performed for these groups using all the
thinking disposition composite and subscales as criterion variables. The results showed no difference between any groups on any of the thinking disposition composite or subscales. In fact none of the ANOVAs even approached significance. Thus these analyses suggest that the thinking dispositions measured in this study do not undergo change throughout higher education.

The analyses above collapse participants across disciplines and thus does not have the resolution power to detect the possible development of thinking dispositions within certain disciplines. To address this issue the above analyses were repeated within three general groupings of academic discipline: science and math, social science, and the arts. Because partitioning the participants into these further groupings significantly reduced sample sizes, participants were not grouped into year enrolled in university, but rather into year 1 & 2 (N = 18, 12, and 4, for science and math, social science and arts respectively) versus year 3 & 4 or more (N = 18, 16, and 16 respectively). Within each of the three academic groupings, a series of t-tests were performed between the year 1 & 2 group versus the year 3 & 4 group. With very few exceptions, the results again indicated that there was no statistical difference in thinking disposition scores (composite or subscale) between those early in their university program versus those late in their program--regardless of academic grouping. The few exceptions were: Students late in their science and math program obtained significantly higher scores than those early in that program in the superstitious thinking subscale $t(34) = 2.659, p < .05$; and students late in their arts program obtained both significantly higher scores than those early in that program on the Need For Closure subscale $t(18) = 2.85, p < .05$, and significantly lower scores on the superstitious thinking subscale $t(18) = 3.48, p < .005$.

One final comparison can be made between the three general classes of disciplines outlined above. Are there any differences in thinking dispositions
between participants who come from science and math, social science and arts programs. A series of one way ANOVAs was performed using general academic discipline (i.e., science and math, social science, and arts) as a grouping variable with all the thinking disposition scales (composite and subscales) alternating as the criterion variable. Once again with very few exceptions the results demonstrated that by-and-large participants across these groupings did not differ in the thinking disposition scales. The exceptions were: A one way ANOVA indicated a significant difference between the three academic groupings on the Openness-Ideas facet of the Revised NEO Personality Inventory $F(2,102) = 6.90, p < .005$, subsequent Fisher PLSD post-hoc comparisons revealed that the arts students scored significantly higher on this subscale than either the science and math or social science students; a separate one way ANOVA also indicated a significant difference between the three academic groupings on the Impulsiveness facet of the Revised NEO Personality Inventory $F(2,102) = 4.10, p < .05$, subsequent Fisher PLSD post-hoc comparisons revealed that the social science students scored higher on this subscale than the science and math students.\(^8\)

Performance on the Projection Tasks

The tendency for people to overly project one's own opinions, attitudes, behaviours etc. when predicting those of others was the focus of the projection tasks. There are two particularly relevant issues to be explored. First, in as much as belief bias is also a tendency to project beliefs onto the world, is belief bias also related to the reported tendency to project one's own opinions, attitudes etc. Second, what are the consequences of projection when it is done in an ecologically

---

\(^8\) During the pilot testing of the Belief Identification scale, however, a significant difference was found between scores obtained by graduate students ($N = 16$) and a group of workers who had not attended graduate school ($N = 13$) $t(27) = 2.29, p < .05$. Although other factors may account for this preliminary finding, this result might suggest that the educational range considered above is too narrow to observe any statistically significant differences.
valid environment? Like the ecological height judgment task, the two tasks making up this portion of the research provide participants with ecologically valid environments. The opinion prediction task assessed how much participants projected their own opinion when predicting the opinions of their school peers. The knowledge prediction task assessed how much participants projected their own knowledge when asked to predict the knowledge held by their school peers.

**Individual Differences in Projection**

The traditional method of analyzing the much investigated projection data in a opinion prediction task was to observe at the group level any discrepancy in consensus estimates between the group of participants who endorsed an item and the group of participants who did not endorse the item. The typical finding was that the group endorsing an item gave significantly higher estimates for the number of people endorsing that item than the did the group who did not endorse the item. Thus each of the two groups perceived a consensus with their own opinion. Those who disagreed with statement x were likely to perceive consensus on their opinion while those who agreed with statement x were also likely to perceive a consensus with their opinion. In as far as both mutually exclusive opinions cannot be correct, this projection of opinion onto others has been termed the “false consensus effect”. Table 21 shows that at this between-subjects level of analysis, the false consensus effect was replicated. The left most column of Table 21 reproduces the items which participants were asked to indicate their endorsement/nonendorsement along with an estimation as to what percentage of the others would endorse the item. The next two columns break up the sample for each item on the basis of nonendorsers (column 2) and endorsers (column 3). A series of t-tests (all two tailed) between the nonendorser group mean estimates and the endorser group mean estimates show that on 76% of the items (22 out of
<table>
<thead>
<tr>
<th>Item</th>
<th>Nonendorsers</th>
<th>Endorsers</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am basically optimistic.</td>
<td>54.17</td>
<td>66.63***</td>
<td>85.25</td>
</tr>
<tr>
<td>2. I think I would like the work of a school teacher.</td>
<td>40.03</td>
<td>46.13</td>
<td>52.46</td>
</tr>
<tr>
<td>3. I would rather spend a quiet evening at home than go out to a party.</td>
<td>39.05</td>
<td>42.43</td>
<td>47.54</td>
</tr>
<tr>
<td>4. It often seems that my life has no meaning.</td>
<td>35.73</td>
<td>50.16**</td>
<td>26.23</td>
</tr>
<tr>
<td>5. I have good leadership abilities.</td>
<td>51.36</td>
<td>58.87</td>
<td>81.97</td>
</tr>
<tr>
<td>6. I do not always tell the truth.</td>
<td>58.75</td>
<td>75.01***</td>
<td>83.61</td>
</tr>
<tr>
<td>7. I would like to spend a year in London or Paris.</td>
<td>62.67</td>
<td>79.08***</td>
<td>87.70</td>
</tr>
<tr>
<td>8. My table manners are not quite as good as when I am out in company.</td>
<td>54.74</td>
<td>70.35***</td>
<td>68.85</td>
</tr>
<tr>
<td>9. My emotional needs are basically satisfied.</td>
<td>42.50</td>
<td>57.47***</td>
<td>72.13</td>
</tr>
<tr>
<td>10. Several times a week I feel as if something dreadful is about to happen.</td>
<td>32.65</td>
<td>46.86**</td>
<td>18.03</td>
</tr>
<tr>
<td>11. I am sensitive to the feelings of others.</td>
<td>47.50</td>
<td>63.40</td>
<td>96.72</td>
</tr>
<tr>
<td>12. People should not have to pay taxes for the schools if they do not have children.</td>
<td>42.97</td>
<td>64.27***</td>
<td>27.87</td>
</tr>
<tr>
<td>13. Television is my primary form of entertainment.</td>
<td>42.36</td>
<td>61.42***</td>
<td>19.67</td>
</tr>
<tr>
<td>14. My daily life is full of things that keep me interested.</td>
<td>48.86</td>
<td>63.42***</td>
<td>80.65</td>
</tr>
<tr>
<td>15. Life on other planets will be discovered within 20 years.</td>
<td>32.40</td>
<td>63.36***</td>
<td>36.07</td>
</tr>
<tr>
<td>16. Sometimes I feel like swearing.</td>
<td>71.67</td>
<td>87.59*</td>
<td>95.08</td>
</tr>
<tr>
<td>17. Police should use whatever force is necessary to maintain law and order.</td>
<td>40.61</td>
<td>66.05***</td>
<td>32.79</td>
</tr>
<tr>
<td>18. I regard the right to speak my mind as very important.</td>
<td>--.--</td>
<td>86.03</td>
<td>100.00</td>
</tr>
<tr>
<td>19. I am frequently depressed.</td>
<td>33.55</td>
<td>47.79**</td>
<td>23.77</td>
</tr>
<tr>
<td>20. I used to like it very much when one of my papers was read to the class in school.</td>
<td>43.17</td>
<td>63.26***</td>
<td>70.49</td>
</tr>
</tbody>
</table>

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

Note: N = 123. Table continued on next page.
Table 21 (continued)

<table>
<thead>
<tr>
<th>Item</th>
<th>Nonendorsers</th>
<th>Endorsers</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. I am concerned about how much sugar I eat.</td>
<td>34.06</td>
<td>51.75***</td>
<td>48.36</td>
</tr>
<tr>
<td>22. I usually don't like to talk much unless I am with people I know very well.</td>
<td>52.13</td>
<td>56.88</td>
<td>49.18</td>
</tr>
<tr>
<td>23. Persons convicted of first degree murder should receive the death penalty</td>
<td>39.63</td>
<td>54.74***</td>
<td>31.97</td>
</tr>
<tr>
<td>24. When in a group of people I have trouble thinking of the right things to talk about.</td>
<td>42.99</td>
<td>48.93</td>
<td>36.89</td>
</tr>
<tr>
<td>25. The government should exercise more control over what is shown on television.</td>
<td>30.39</td>
<td>54.00***</td>
<td>42.62</td>
</tr>
<tr>
<td>26. I usually take an active part in the entertainment at parties.</td>
<td>44.39</td>
<td>47.54</td>
<td>57.38</td>
</tr>
<tr>
<td>27. I am a well organized person.</td>
<td>45.86</td>
<td>54.80*</td>
<td>70.49</td>
</tr>
<tr>
<td>28. I cannot keep my mind on one thing.</td>
<td>45.76</td>
<td>55.13**</td>
<td>51.64</td>
</tr>
<tr>
<td>29. I enjoy hearing lectures on world affairs.</td>
<td>33.83</td>
<td>47.68***</td>
<td>66.39</td>
</tr>
<tr>
<td>30. I refuse to play some games because I am not good at them.</td>
<td>48.00</td>
<td>55.10*</td>
<td>48.36</td>
</tr>
</tbody>
</table>

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

Note: N = 123.

29 items, item 18 is excluded since their were no nonendorsers of this item.

Participants who endorsed an item gave significantly higher estimates for the proportion of others who would also endorse that item as compared to the nonendorser group. The last column of the table displays the actual percentage of endorsement for the item.

Table 22 is completely analogous in structure to Table 21. Table 22 displays an identical between-subjects level of analysis on the knowledge prediction task. Rather than two groups formed on the basis of item endorsement, the two groups in the knowledge prediction task were formed as a function of whether they answered the item correctly or not. On 83% of these knowledge questions, significantly higher estimates for the number of participants who answer correctly were given by the group who in fact correctly answered the item as opposed to the estimates of the group who answered incorrectly. Again a false
Table 22
Mean Percentage Estimate of Correct Response for Each Item on the Knowledge Prediction Task by Subjects Who Incorrectly Answered and Correctly Answered the Item

<table>
<thead>
<tr>
<th>Item</th>
<th>Incorrectly Answered</th>
<th>Correctly Answered</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What is the name of the largest ocean on Earth?</td>
<td>65.22</td>
<td>81.841***</td>
<td>67.83</td>
</tr>
<tr>
<td>2. What was the last name of the man who was the radio broadcaster for the &quot;War of the Worlds&quot;?</td>
<td>34.71</td>
<td>58.07***</td>
<td>26.96</td>
</tr>
<tr>
<td>3. What is the name of the actor who portrayed Sergeant Friday on &quot;Dragnet&quot;?</td>
<td>32.87</td>
<td>30.00</td>
<td>0.80</td>
</tr>
<tr>
<td>4. What is the only liquid metal at room temperature?</td>
<td>51.84</td>
<td>69.85***</td>
<td>73.04</td>
</tr>
<tr>
<td>5. What Italian city was destroyed when Mount Vesuvius erupted in 79 A.D.?</td>
<td>40.21</td>
<td>58.26***</td>
<td>46.96</td>
</tr>
<tr>
<td>6. What is the name of the first artificial satellite put in orbit by Russia in 1957?</td>
<td>41.73</td>
<td>57.41***</td>
<td>55.65</td>
</tr>
<tr>
<td>7. What was the name of the supposedly unsinkable ship that sunk on its maiden voyage in 1912?</td>
<td>33.14</td>
<td>89.98***</td>
<td>93.91</td>
</tr>
<tr>
<td>8. What is the last name of the man who began the reformation in Germany?</td>
<td>42.86</td>
<td>56.58***</td>
<td>33.04</td>
</tr>
<tr>
<td>9. What is the name of the inventor of the Steamboat &quot;Clermont&quot;?</td>
<td>27.69</td>
<td>12.5 (n=2)</td>
<td>1.70</td>
</tr>
<tr>
<td>10. What was the last name of the Ventriloquist who provided the voice for Charlie McCarthy?</td>
<td>21.36</td>
<td>44.29***</td>
<td>6.09</td>
</tr>
<tr>
<td>11. What is the name for a Medical Doctor who specializes in diseases of the skin?</td>
<td>49.15</td>
<td>75.07***</td>
<td>61.29</td>
</tr>
<tr>
<td>12. Who was the Egyptian queen who joined forces with Mark Anthony of Rome?</td>
<td>45.38</td>
<td>74.56***</td>
<td>86.09</td>
</tr>
<tr>
<td>13. What is the name of the largest desert on Earth?</td>
<td>49.75</td>
<td>78.49***</td>
<td>86.09</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001, all two-tailed.
Note: N = 115. Table continued on next page.
Table 22 (continued)

<table>
<thead>
<tr>
<th>Item</th>
<th>Incorrectly Answered</th>
<th>Correctly Answered</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>14. What is the Capital of Denmark?</td>
<td>55.28</td>
<td>62.66</td>
<td>35.65</td>
</tr>
<tr>
<td>15. What is the name of the unit of measure that refers to a six-foot depth of water?</td>
<td>40.48</td>
<td>48.70</td>
<td>20.00</td>
</tr>
<tr>
<td>16. Of which country is Buenos Aires the Capital?</td>
<td>52.55</td>
<td>61.86*</td>
<td>44.34</td>
</tr>
<tr>
<td>17. What is the unit of electrical power that refers to a current of one Ampere at one Volt?</td>
<td>50.52</td>
<td>62.76**</td>
<td>33.04</td>
</tr>
<tr>
<td>18. What is the name of the process by which plants make their food?</td>
<td>62.78</td>
<td>77.93**</td>
<td>92.17</td>
</tr>
<tr>
<td>19. What is the last name of the man who first studied genetic inheritance in plants?</td>
<td>40.36</td>
<td>59.4***</td>
<td>39.13</td>
</tr>
<tr>
<td>20. What is the name of the North Star?</td>
<td>45.29</td>
<td>49.89</td>
<td>24.35</td>
</tr>
<tr>
<td>21. What is the name of the singer who popularized a dance known as the &quot;Twist&quot;?</td>
<td>43.56</td>
<td>68.18***</td>
<td>44.35</td>
</tr>
<tr>
<td>22. What was the name of the Zeppelin that exploded in Lakehurst N.J. in 1937?</td>
<td>25.2</td>
<td>57.6***</td>
<td>21.74</td>
</tr>
<tr>
<td>23. Andy Griffith was the Sheriff of what town on television's &quot;Andy Griffith Show&quot;?</td>
<td>38.69</td>
<td>60.54***</td>
<td>32.17</td>
</tr>
<tr>
<td>24. What is the name of the collar bone?</td>
<td>41.92</td>
<td>57.90***</td>
<td>26.96</td>
</tr>
<tr>
<td>25. What is the longest river in South America?</td>
<td>41.33</td>
<td>75.78***</td>
<td>46.96</td>
</tr>
<tr>
<td>26. What animal runs the fastest?</td>
<td>58.93</td>
<td>77.86***</td>
<td>75.65</td>
</tr>
<tr>
<td>27. For which country is the Yen the monetary unit?</td>
<td>62.59</td>
<td>83.36***</td>
<td>74.78</td>
</tr>
<tr>
<td>28. What is the name of the river on which Bonn is located?</td>
<td>37.89</td>
<td>60.41***</td>
<td>29.57</td>
</tr>
<tr>
<td>29. What is the name of the lizard that changes its colour to match the surroundings?</td>
<td>41.59</td>
<td>84.66***</td>
<td>81.70</td>
</tr>
<tr>
<td>30. What is the last name of the famous magician and escape artist who died of appendicitis?</td>
<td>38.33</td>
<td>73.38***</td>
<td>76.52</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001  
Note: N = 115.
Consensus-like effect is apparent at this between-subjects level of analysis.

Thus on both of the projection tasks, the traditional between-subjects analysis demonstrates that participants are likely to project their opinion and knowledge on to others. Although the false consensus effect is a replication of much research on the opinion prediction task, it’s analog in the projection of knowledge has not been previously demonstrated.

The traditional between-subjects analyses shown above fails to address at least two important issues. First, it collapses across participants and thus provides no information concerning individual differences—a primary concern of the present research, and secondly, it fails to investigate the consequences of projection on predictive accuracy (Dawes, 1989, 1990; Hoch, 1987; Krueger & Zeigler, 1993). Recall the results of the ecological height judgment task in which the projection of the gender/height relationship actually led to more accurate height estimates. This was the case since in the real world, gender is correlated with height, i.e., gender is a diagnostic cue. Thus if one’s opinion or knowledge is diagnostic of the population they are predicting, then it may very well be efficacious to project it. Couple this with the truism that most people are in the majority, you then have a situation where indeed the held opinion, attitude, behavior etc. should be diagnostic for most people.

Using a different type of analysis, Hoch (1987), effectively demonstrated that the observed projection of opinion, previously shown to lead to a false consensus effect in between-subjects level of analyses, may actually be efficacious. Hoch utilized a Brunswikian analysis focusing on the performance of individual participants aggregated across items rather than the traditional strategy of considering the performance on individual items aggregated across participants. Hoch’s analysis begins with defining some parameters: For any given item, (T) represents the target position—the actual percentage agreement in the target
population, (O) represents own position—the participant’s own agreement with the item (scored 0/1) and (P) represents the prediction—the subject’s prediction of the percentage of agreement in the population. To accommodate the knowledge projection task used in this study, these parameters will denote the functionally analogous following parameters: (T) will denote the actual percentage of correctly answering an item on the knowledge prediction task, (O) denotes the participant’s own performance on the item (scored 0/1 for incorrect/correct answer) and (P) represents the subject’s prediction of the percentage of correct performance in the population.

Thus one possible index of predictive accuracy is the correlation, across items, between the actual percentage agreement in the target population and the participant’s prediction of the level of agreement in the population, i.e., r(T,P). This index is identical in structure to r(T,E)—the correlation between target height (T) and participants estimates (E) across stimuli—accuracy index employed for the height judgment tasks. The r(T,P) index was calculated for each participant for both projection tasks. The mean r(T,P)_opinion for the 122 participants completing the opinion prediction task was .492 (SD = .224), the mean r(T,P)_knowledge for the 115 participants completing the knowledge prediction task was .646 (SD = .135). Participants were thus found to exhibit moderate to high predictive accuracy for both of the projection tasks (p < .001 for both correlations).

As already outlined above, most participants will have positive correlations between their own opinion/performance (O) and the target position (T). This was indeed found to be the case in the present investigation in which the mean r(T,O)_opinion correlation was .534 (SD = .206), and the mean r(T,O)_knowledge correlation was .584 (SD = .132). These two sets of correlations suggest that participants in the opinion prediction task could have been slightly more accurate had they projected their own opinion/performance even more i.e., r(T,O) > r(T,P).
Interestingly, this was not the case for the knowledge prediction task. Had participants projected any more than they already had in the knowledge prediction task, their predictive accuracy would have decreased, i.e., \( r(T,O) < r(T,P) \). To be more specific, other information that participants based their predictions on in the knowledge prediction task proved of higher diagnostic value than their own position (O). Just how much participants did project is indexed by the correlation between prediction and own position i.e., \( r(P,O) \) viz., an index of perceived consensus. Participants clearly perceived consensus on the projection tasks. The mean \( r(P,O)_\text{opinion} \) was .517 (SD = .213), and the mean \( r(P,O)_\text{knowledge} \) was .653 (SD = .126). Thus participants perceived less consensus on the opinion prediction test than their really was, i.e., \( r(P,O)_\text{opinion} < r(T,O)_\text{opinion} \). This sort of result suggests that “false consensus” may be a misnomer. Rather than being false, as suggested by a between-subjects analysis, the projection of consensus often improves predictive accuracy. Participants in the knowledge prediction task, however, perceived more consensus than there really was, i.e., \( r(O,P)_\text{knowledge} > r(T,O)_\text{knowledge} \). Thus in the knowledge prediction task, participants may have projected too much.

The existence of a correlation between own position and target position for most participants in these samples meant that predictive accuracy was at least to some point dependent on the projection of one’s own opinions/knowledge. That predictive accuracy is at least partially dependent on some projection does not rule out that too much projection will also decrease predictive accuracy. The consequence of perceived consensus on predictive accuracy for each task can be indexed by taking the correlation between the degree of perceived consensus \([r(P,O)] \) and the resulting predictive accuracy \([r(T,P)] \), i.e., \( r[r(P,O),r(T,P)] \). The \( r[r(P,O),r(T,P)]_\text{opinion} \) was .443. This correlation is of moderate size and again suggests that projection in the opinion prediction task resulted in improved
accuracy. Once again, this was not necessarily the case with the knowledge prediction task. The resulting $r[r(P,O)\cdot r(T,P)]_{\text{knowledge}}$ correlation was a nonsignificant .125.

These results are also consistent with an analysis using a parameter introduced by Krueger and Zeiger (1993). They suggested that an index of projection at the individual level that is relative to the predictive accuracy attained can be derived by correlating for each subject, across items, own position ($O$, scored as 0/1) with the difference between the actual target percentage and the predicted target percentage i.e., $T - P$. While the mean correlation of $r(O, P-T)_{\text{opinion}}$ was -.060 (SD = .231)–which was significantly less than zero ($t(121) = -2.86, p < .005$) the mean correlation of $r(O, P-T)_{\text{knowledge}}$ was -.040 (SD = .256)–which was not significantly different from zero ($t(114) = -1.67, \text{ns}$). Thus, as a whole, the sample of participants in the opinion prediction task displayed a statistically significant tendency not to over project opinion, but rather to under project opinion—what Krueger and Zeiger termed a “false uniqueness effect”. This was not the case in the knowledge prediction task. A tendency to either over project (false consensus) or to under project (false uniqueness) was not statistically significant in the knowledge prediction task.

The preceding results concerning the opinion prediction task converge with those critiques of the traditional “false consensus” interpretation of projection (Dawes, 1989, 1990; Hoch, 1987; Krueger & Zeigler, 1993; also see Stanovich and West, 1998a). Various indices showed that projection facilitated predictive accuracy, and once more, predictive accuracy actually could have been improved for the sample as a whole if participants had projected even more. The results concerning the knowledge prediction task are less clear. In contrast to the opinion prediction task, there was some indication that participants actually did project more knowledge than was warranted. Unlike the opinion prediction task,
it was also evident that participants in the knowledge prediction task would have been less accurate had they projected any more than they already had. It would appear that other diagnostic information employed by participants in the knowledge prediction task allowed them to attain a level of predictive accuracy not possible if they had simply projected their own knowledge. The opposite was true with the opinion prediction task—participants would have done better simply projecting opinion than employing what ever cues they had.

Direct measures of the degree of projection participants exhibited in these social perception tasks (i.e., the $r(O,P)$ indices) was not statistically related to any performance measures on the prior belief tasks. The $r(O,P)_{\text{opinion}}$ index, however, did correlate with both general cognitive ability 1 ($r = .181, p < .05$), and general cognitive ability 2 ($r = .183, p < .05$)—both small but statistically significant. Thus not only was it the case that those participants who projected consensus in the opinion prediction task attained higher levels of predictive accuracy, they also tended to score higher on measures of cognitive ability. The equivalent projection index for the knowledge prediction task was not correlated with cognitive ability. These two results are particularly interesting and perhaps anticipated from an adaptive view of intelligence. Once again higher levels of projection were correlated with cognitive ability only when doing so lead to predictive accuracy. A very different pattern of covariation with the prior belief tasks emerged when the predictive accuracy measures of the projection tasks were analyzed.

Table 23 displays the zero-order correlations between the predictive accuracy measures of the projection tasks (i.e., the $r(T,P)$ indices) with various indices of performance on the prior belief tasks. Performance on both projection tasks was significantly correlated with performance on the syllogisms test, the ability to inhibit prior belief when it would be efficacious to do so on a perceptual judgment
task (Height Adjustment variable), the ability to evaluate verbal arguments (AET) and covariation data (covariation judgment) independently of prior beliefs. Performance on both projection tasks was also positively correlated with both cognitive ability measures and the AOT composite.

Table 23
Zero-Order Correlations Between the r(T,P) Predictive Accuracy Indices of the Projection Tasks with Performance Indices of the Prior Belief Tasks, Cognitive Ability Measures and the Actively Open-Minded Thinking Scale (AOT) Composite

<table>
<thead>
<tr>
<th>Variable</th>
<th>r(T,P) Opinion</th>
<th>r(T,P) Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllogisms, Belief Bias</td>
<td>-.283**</td>
<td>-.357***</td>
</tr>
<tr>
<td>r(T,E) Ecological†</td>
<td>.175</td>
<td>.186</td>
</tr>
<tr>
<td>r(G,E) Ecological†</td>
<td>.221*</td>
<td>.127</td>
</tr>
<tr>
<td>r(T,E) Matched†</td>
<td>-.017</td>
<td>-.093</td>
</tr>
<tr>
<td>r(G,E) Matched†</td>
<td>-.028</td>
<td>-.127</td>
</tr>
<tr>
<td>Height Adjustment†</td>
<td>.219*</td>
<td>.257**</td>
</tr>
<tr>
<td>AET</td>
<td>.260**</td>
<td>.327***</td>
</tr>
<tr>
<td>Covariation Judgment‡‡</td>
<td>.265**</td>
<td>.254**</td>
</tr>
<tr>
<td>Confounded Variable‡‡‡</td>
<td>-.067</td>
<td>-.135</td>
</tr>
<tr>
<td>Cognitive Ability 1</td>
<td>.332***</td>
<td>.489***</td>
</tr>
<tr>
<td>Cognitive Ability 2</td>
<td>.293**</td>
<td>.354***</td>
</tr>
<tr>
<td>AOT Composite</td>
<td>.416***</td>
<td>.354***</td>
</tr>
</tbody>
</table>

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

Hierarchical regression analyses was performed to determine if the AOT composite accounted for any additional variance in predictive accuracy once all the variance attributable to cognitive ability is partialled out. The first analysis on Table 24 shows that the AOT composite accounted for a significant portion of
Table 24
Hierarchical Regression Analyses Predicting Predictive Accuracy on the Projection Tasks with Cognitive Capacity Measures Entered First Followed by The Actively Open-minded Thinking Scale (AOT)

r(T,P) Opinion Prediction
(N = 123)

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>$R^2$ Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cognitive Ability 1</td>
<td>.332</td>
<td>.110</td>
<td>14.85***</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.341</td>
<td>.006</td>
<td>0.83</td>
<td>.083</td>
</tr>
<tr>
<td>3.</td>
<td>AOT</td>
<td>.461</td>
<td>.097</td>
<td>14.44***</td>
<td>.330</td>
</tr>
</tbody>
</table>

r(T,P) Knowledge Prediction
(N = 115)

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>R</th>
<th>$R^2$ Change</th>
<th>F Change</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cognitive Ability 1</td>
<td>.489</td>
<td>.239</td>
<td>35.56***</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Cognitive Ability 2</td>
<td>.490</td>
<td>.001</td>
<td>0.04</td>
<td>.019</td>
</tr>
<tr>
<td>3.</td>
<td>AOT</td>
<td>.524</td>
<td>.035</td>
<td>5.35*</td>
<td>.214</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01, *** = p < .001

additional variance in the opinion prediction even after both cognitive ability measures were entered into the regression first. In fact the AOT composite seems to be a more potent predictor of predictive accuracy on the opinion prediction task than both cognitive ability measures combined. The top commonality analysis displayed in Table 25 shows that the AOT composite accounted for an appreciably larger portion of unique variance (12%) than did both cognitive ability measures combined (4.5%). The bottom hierarchical analysis in Table 24 also shows that the AOT composite accounted for additional variance in the knowledge prediction task once all the variance attributable to cognitive ability is partialled out. The second commonality analysis in Table 25
shows that while cognitive ability seemed to account for an appreciably larger portion of unique variance in the predictive performance on the knowledge prediction task, there was a substantial amount of common variance explained by both cognitive ability and the thinking disposition scale.

Table 25
Commonality Analyses of Thinking Disposition Measures and Cognitive Ability Measures on the Predictive Accuracy Indices of The Opinion Prediction and Knowledge Prediction Tasks

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Unique</th>
<th>Common</th>
<th>Total Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>r(T,P) Opinion Prediction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOT</td>
<td>.120</td>
<td>.296</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.045</td>
<td>.296</td>
<td></td>
</tr>
<tr>
<td><strong>r(T,P) Knowledge Prediction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOT</td>
<td>.033</td>
<td>.321</td>
<td></td>
</tr>
<tr>
<td>Cognitive Ability</td>
<td>.170</td>
<td>.321</td>
<td></td>
</tr>
</tbody>
</table>

Note: Cognitive Ability = variance accounted for by both Cognitive Ability 1 and Cognitive Ability 2 composite scales.

Thus it was those participants who were more successful on the prior belief tasks that also exhibited greater success at tracking others' opinions or knowledge. This was especially true with their ability to properly track the knowledge of others. Cognitive ability was once again positively correlated with success on these projection tasks--again consistent with the notion of intelligence leading to effectiveness across domains. These results are anticipated by Baron (1985) "intelligence is the set of properties within a psychological theory that make for effectiveness, regardless of the environment a person is in" (p. 15; see also
Belief Bias

Anderson, 1990, 1991; Larrick, Nisbett & Morgan, 1993; Oaksford & Chater, 1993, 1994, 1995; Schoemaker, 1991, Shanks, 1995). Even more striking was the significant contribution to this predictive accuracy by actively open-minded thinking dispositions. This is particularly notable in the opinion prediction task where the AOT composite actually accounted for substantially more unique variance in predictive accuracy than both cognitive ability measures combined.
Summary and Conclusions

The results of this study showed large individual differences across all the prior belief and social perception (projection) tasks. These individual differences, furthermore, yielded significant associations of relevance to all three organizational themes of the present research. A review of these results and their implications for the issues brought up in the introduction are addressed below.

Belief Bias and the Normative Issue

Belief bias was interpreted here to represent a nonnormative tendency in human reasoning. As such, it was hypothesized that individuals exhibiting higher levels of cognitive ability (as indexed by traditional tests of psychometric intelligence) would be both less likely to exhibit belief bias (when prior belief was nondiagnostic) and more successful at reasoning independently of prior beliefs in comparison to participants demonstrating lower levels of cognitive ability. The results were clearly consistent with this prediction. Participants with higher levels of cognitive ability were indeed more likely to give the normative response for each of the prior belief tasks. The series of median split analyses performed on the non-height-judgment-prior-belief tasks (summarized in Table 1) demonstrate that participants giving or approximating the normative response scored significantly higher on measures of cognitive ability. This finding is also evident in a separate correlational analysis (see Table 3) where again normative performance on all the prior belief tasks was found to correlate significantly with the cognitive ability measures.

The median split analyses performed on each of the four non-height judgment prior belief tasks also demonstrated that participants giving or approximating the normative response on each task were also different from the group not approximating the normative response on several other variables of
relevance to the normative issue. On two of the four tasks, participants giving the normative response also reported significantly higher high school or university averages, scored significantly higher on three of the four tasks on at least one of the two print exposure indices, and scored higher on a test of general knowledge in three of the four tasks (see Table 1). In no case did participants giving the nonnormative response score higher on any of these variables. In addition to this, with the exception of the confounded variable task9, participants giving the normative response on any one of the prior belief tasks were also more likely to give the normative response on another prior belief task (see Table 3). This pattern of results was also evident in the social perception tasks used in this research. Predictive accuracy on both the opinion prediction task and knowledge prediction task was related to normative performance on the prior belief tasks and cognitive ability (see Table 23). One other important implication of this is the fact that normative performance on tasks in which the status of the normative model might be questionable is correlated with normative performance in tasks in which there is no dispute about the status of the appropriate normative model.

What emerges in this pattern of data is not only that the same participants by-in-large consistently gave or approximated the normative (or optimal) response across these diverse prior belief tasks, but also that these participants are differentiated from those participants who do not approximate the normative response by significantly higher levels of cognitive ability, generally higher levels of academic achievement, greater levels of exposure to print material, and also being generally more knowledgeable. Thus the descriptive data showed that normative responses were consistently endorsed by those participants who would be expected to be more effective reasoners. A critic of the normative models used here as benchmarks of cognitive performance in these prior belief tasks would

---

9 Although correlating with both cognitive ability measures, performance on the confounded variable task only correlated with AET performance (p < .05) and was not significantly related to any other prior belief task.
find this result difficult to reconcile with any adaptive view of cognition.

The normative response on the above reasoning tasks requires that prior beliefs not be projected on to the evaluation of evidence. For the most part, these tasks are designed such that prior beliefs conflict with the optimal response. This of course was not the case in the ecological height judgment task. Unlike the matched height judgment task, prior belief in the ecological task accurately reflects an aspect of the perceptual environment. In contrast, prior belief in the matched task is inconsistent with the relationships in the stimulus set—a situation analogous to the logic of the other prior belief tasks. Due to the fact that the prior belief concerning the gender/height relationship remained accurate, projection of belief in the ecological height judgment task was actually found to facilitate rather than hamper height judgment accuracy. Furthermore, the same cognitive ability measures which displayed moderate negative correlations with belief bias on the verbal and correlational reasoning tasks also correlated negatively with belief bias in the matched height judgment task, and correlated positively with belief bias on the ecological task. The difference in these dependent correlations was statistically significant ($p < .01$). Thus, cognitive ability was positively correlated with belief bias in a perceptual judgment task when prior belief accurately reflected an aspect of the perceptual environment and negatively correlated with belief bias in a perceptual judgment task when prior belief inaccurately reflected an aspect of the perceptual environment—just the pattern that an adaptive view of intelligence would predict (Anderson, 1990, 1991; Larrick et al., 1993). An analogous result was also evident in the social perception tasks. Higher levels of projection were significantly correlated with cognitive ability only when increased projection would lead to higher predictive accuracy (as was the case in the opinion prediction task).

Another series of median split analyses also produced results that were
consistent with the idea that cognitive ability leads to effectiveness across environments. Participants projecting belief in the ecological height judgment task were not only differentiated from those participants not projecting belief by higher levels of cognitive ability, but they also showed a clear trend (although not statistically significant) of having been exposed to more print ($p = .071 \text{ to } .076$) and another trend of reporting higher high school averages ($p = .078$). This pattern, however, once again reverted to that observed in the above prior belief tasks when it becomes inefficacious to project belief. Median splits performed on the ability to avoid projecting the gender/height relationship in the matched height judgment task (factoring in the magnitude of the prior belief) clearly showed that it was those participants with higher levels of cognitive ability that were more successful at inhibiting the prior belief cue from influencing their judgment. Participants who did not project belief also scored significantly higher on both print exposure measures as well as the test of general knowledge.

The Domain Specificity and Generality of Belief Bias

The results of this study were also consistent with the hypothesis that belief bias would exhibit some domain generality. With the exception of the confounded variable task, performance on all the prior belief tasks were found to significantly correlate with one another (see Table 3). The significant correlation between the AET and syllogistic reasoning task is perhaps not surprising given that these two tasks are both ultimately vested in verbal reasoning. Nevertheless, a performance measure on a task of pure deductive logic was found to correlate with a performance measure of informal argument reasoning. The covariation judgment task, which requires some level of numerical reasoning, was also correlated with both of these verbal reasoning tasks. This is an interesting result given that the nature of the covariation task is notably different from the other two verbal reasoning tasks. Much more surprising, however, is that the magnitude
of the correlations involving performance on the matched height judgment task were just as large. Of particular note is the result from a direct test of the generalizability of belief bias: The ability to decouple the influence of prior knowledge from deductive reasoning was associated with the ability to decouple the influence of prior knowledge on perceptual judgments. This occurred despite the fact that the height judgment task involved no verbal reasoning and was a purely perceptual task. The correlations observed between the height adjustment score and the syllogistic reasoning task, AET, and covariation judgment task do not simply appear to be due to their sharing of overlapping verbal reasoning requirements. This resulting uniformity of the correlations across tasks with distinctively different processing demands suggests that there is some true domain generality in the ability to reason about the world in the face of interfering prior belief. These observed correlations between the prior belief tasks, however, only succeed in refuting an extreme position of domain specificity. The magnitude of these intercorrelations still leaves much room for appreciable domain specificity in the underlying cognitive mechanisms responsible for biased reasoning.

A series of hierarchical regression analyses revealed that while some of this domain generality was explained by an underlying commonality in cognitive ability resources, various indices of performance across these tasks continued to be significantly associated with one another even after partialling out all the variance attributable to cognitive ability. The only exception to this was the association between performance on the AET and the covariation judgment task which did seem to be a clear instance in which their correlation may indeed have resulted from their common reliance on cognitive ability.

The Utility of the Cognitive Capacity - Thinking Dispositions Framework in the Prediction of Belief Bias

With one minor exception, performance on all the prior belief tasks was
significantly correlated with the cognitive ability measures (see Table 3). This was even the case for the confounded variable task which with the exception of the AET, failed to correlate with the other prior belief tasks. On each task requiring participants to decouple prior belief influence from reasoning and judgment, it was clearly those participants with higher levels of cognitive ability that were most successful at achieving this goal. The pattern of results involving the cognitive ability measures and all the prior belief tasks were once again consistent with an adaptionist view of intelligence (Anderson, 1990, 1991; Baron, 1985, pp. 15; Larrick et al., 1993). Cognitive ability correlated with belief biased judgments when the focal belief accurately reflected an aspect of the task environment (i.e., the ecological height judgment task) and negatively correlated with belief bias when the focal belief was knowingly incongruent with the task environment (i.e., all the remaining prior belief tasks). Cognitive ability alone, however, was not the only factor found to account for performance variation. The results were also consistent with the hypothesis that thinking disposition measures will also predict performance, and some mild to moderate support for the hypothesis that thinking dispositions would predict additional variation in task performance once all the variance accounted for by cognitive ability is partialled out.

The construction of the actively open-minded thinking (AOT) composite was largely informed by the results of a principal components analysis. The AOT composite was found to correlate with performance on the syllogisms test, AET, and covariation judgment task. The AOT composite scale correlated particularly well with the beta weight for argument quality on the AET (r = .436, p < .001).

---

10 Performance on the ecological height judgment task—a task notably different from the other prior belief tasks in that prior belief was not invalidated or pitted against optimal response—approached but did not reach significance with the Cognitive Ability 1 measure. It did, however, significantly correlate with the Cognitive Ability 2 measure.

11 This was also true for the social perception task in which projection lead to facilitating accuracy, i.e., the opinion prediction task.
That this was the case should not come as a great surprise. Evaluating arguments on the AET would seem to clearly engage participants in an activity that is very much influenced by held epistemistic goals and values. Whether an individual participant directs their efforts to justify why a belief-incongruent argument must be poor, for example, or conversely, approaches this belief-incongruent argument with some concern for being an objective processor of information is precisely the type of epistemic goals and values that the AOT composite was designed to tap.

The AOT composite scale, furthermore, continued to account for additional variance on the AET after partialling out all the variance accounted for by both cognitive ability measures. Both the conscientiousness and superstitious thinking composite scales also accounted for additional variance on AET performance after partialling out all the variance attributable to cognitive ability first. The combination of the three thinking disposition scale composites actually accounted for more unique variance on the AET than both cognitive ability measures combined (see Table 14). Thus the results suggest that one can predict performance on argument evaluation more accurately from gauging thinking dispositions than is possible via the use of measures obtained from psychometric tests of intelligence.

Although small in magnitude, the AOT composite scale achieved some level of success at accounting for additional variance in the belief bias index for the syllogisms test once all the variance accounted for by cognitive ability measures is partialled out first. The relation, however, was not statistically significant at the $p < .05$ level (see Table 17). The AOT composite, however, was more successful with other performance indices of the syllogism test. Specifically, the additional variance accounted for by the AOT composite in the inconsistent syllogism score

---

12 The conscientiousness and superstitious thinking scales were not, however, found to account for any additional variance above that already attributable to cognitive ability on any other prior belief tasks and will not be discussed any further.
(a score analogous in structure to the AET index) was moderately better than was
the case for the belief bias index, and did reach statistical significance at the $p < .01$ level.

As cautioned earlier, it must be kept in mind that all of these analyses
represent worst case scenarios. All the variance in task performance that was
shared by both cognitive ability and thinking dispositions was attributed to the
former. A series of commonality analyses displayed on Table 20 puts this issue
into some perspective. Thus, for example, while the AOT composite explained a
significant 4.2% of unique variance in the syllogism inconsistent score, it also
accounted for a further 13.6% of the variance that was in common with cognitive
ability.

Once all the variance accounted for by cognitive ability is partialled out, the
AOT composite was not found to account for any additional variance on the
remaining prior belief tasks. This brings up an important issue in terms of the
utility of thinking disposition measures. The hypothesis that thinking disposition
measures may account for performance on reasoning tasks is partially supported
by the results. It seems, however, that the utility of thinking dispositions lending
to actively open-minded thinking may very well be moderated by considerations of
the types of reasoning tasks--or more broadly the nature of the required
reasoning. While the AOT composite scale was successful in accounting for
additional variance on the AET, and mildly successful at accounting for
additional variance on syllogistic reasoning, it added nothing to the prediction of
performance on the covariation judgment task, the confounded variable task and
the height judgment tasks--it failed to even attain significant zero order
correlations with the latter two.

A possible explanation for this pattern of results begins with the
consideration that the AOT composite scale is designed to tap an individuals
epistemic goals and values, i.e., it deals largely with epistemic rationality. A willingness to consider conflicting evidence or alternative opinions, to engage in counterfactual reasoning and not to be intimately vested in the continuation of held beliefs are characteristics of relevance to epistemic rationality. The AET presents arguments in which their strengths or weaknesses have direct relevance to the respective social or political issue in the real world. A good argument on the AET about some issue on the test is just as equally a good argument about that issue in the real world. Thus the relevance of the AET arguments to the real world survives any artificial aspects of the test situation. This is not true to the same extent with the other prior belief tasks. Participants in the height judgment tasks, for example, are presented with a task at the other extreme. The stimuli used in the height judgment tasks need not have any relevance at all to a participant’s beliefs about the world. The matched version of that task actually makes the artificial nature of the stimuli explicit in the instructions. Recall that participants in the matched height judgment task are told about the matching process used to equate the male and female heights in the stimuli. Similarly, both the covariation judgment and confounded variable task can easily be construed by participants as employing fictitious and arbitrary numerical data. Thus the covariations assessed in both these tasks need not—and probably do not—survive the test situation and exert some real impact on the participant’s belief system about the world. Syllogistic items may also be seen to have some degree of this arbitrary or artificial aspect. Consider the following syllogism (item 16 on syllogisms test): Premise 1: All mammals walk. Premise 2: Whales are mammals. Conclusion: Whales walk. A total of 90.3% of the participants answered correctly that this syllogism is logically valid. It would be fair to state that none of these participants actually believe whales walk. What leads them to the correct answer is in part an appreciation for the artificiality of
the argument. But the very allowance given to the artificial nature of the argument also serves to shield the participant's belief system from any implications the argument may have for their belief system. That the AET may also be approached as somewhat artificial in construction might be argued. The salience of this artificiality and arbitrariness, however, seems significantly larger for the syllogism arguments. An additional factor that clearly distinguishes informal argument evaluation from syllogistic reasoning pertains to their conduciveness to different sorts of reasoning errors.

Wilson and Brekke (1994) outlined an important distinction in reasoning and decision errors. The first type of reasoning error is the failure to apply the appropriate rule or algorithm to the problem at hand. The second type of error outlined by Wilson and Brekke (1994) occurs in situations where there exists no appropriate rule or algorithm for the problem. This second situation is conducive to unwanted influences on judgments and evaluations—what Wilson and Brekke call mental contamination. An example of such mental contamination is the overestimation of what could have been anticipated in foresight once the outcome is known—viz., hindsight bias (e.g., Fischhoff, 1975, 1982; see also Mitchell, Robinson, Isaacs, & Nye, 1996). Whereas an incorrect evaluation of a syllogistic item can be viewed as a failure to properly apply a rule of logic, there is no clear-cut rule or algorithm to apply to the AET items. Thus the AET is susceptible to “mental contamination”—via prior beliefs—in a manner not possible in the syllogisms test. Mental contamination marks a further distinction between these two verbal reasoning tasks that may have relevance to the cognitive capacity-thinking disposition framework. Cognitive capacity may be particularly functional in the failures to apply appropriate rules or algorithms as computational limitations clearly limit which rules, algorithms or mental operations can be executed by a cognitive system. Although thinking dispositions
probably also influence the selection of the rules/algorithms executed by the
cognitive system, thinking dispositions may be particularly functional for
guarding against unwanted influences on reasoning and judgment (in this case,
from prior beliefs).

**Limitations of the Present Study**

As seems to be true with much research in psychology, the generalizability of
the present study may be criticized due to a restrictive selection of the
participants--virtually all of whom were drawn from a university student
population. One such consequence of not including a wider range of participants
is the over-representation of participants having above average cognitive ability.
A total of 66% of the participants in the current study had pro-rated IQ scores over
the population average of 100 (as calculated from the administered WAIS-R short-
form using Sattler's 1988 formulas). Although this may pose a generalizability
problem, at least one important result indicates that inclusion of more lower
cognitively able participants would have actually boosted the potency of the
thinking disposition measures. A median split was performed on the general
cognitive ability 1 variable resulting in a low cognitive ability group and a high
cognitive ability group. A correlation analysis between the actively open-minded
thinking scale and the beta for argument quality on the AET was performed
separately for both cognitive ability groups. Whereas the correlation between the
AOT composite and the AET index for the low cognitive ability group was .511 ($p <
.0001$), the same correlational relationship for the high cognitive ability group was
.198, ns. Thus the actively open-minded thinking dispositions were
disproportionately beneficial to participants in the lower-end of cognitive ability--
Another limitation of the study was the small number of items on several of the tasks that were used to calculate pertinent variables. The prior belief tasks used in this study are not psychometric instruments purposefully designed to elucidate individual differences. Although the reliability of these measures may be critiqued on this issue, this aspect of the study also serves to underestimate the actual covariance existing between the prior belief tasks. The difference score for belief bias in syllogistic reasoning was calculated using only 16 items, the other major index for that task (inconsistent score) was calculated from only 8 items, the argument quality beta in the AET was calculated for each participant from only 23 data points, the beta weight for $\Delta p$ on the covariation task was calculated for each participant from only 25 data points—and despite all this, these variables still displayed significant intercorrelations with one another as well as with measures of cognitive ability and thinking dispositions. Thus the extent of the generalizability of belief bias may very well have been underestimated in the current study.

Another limitation is the short-coming of some of the tasks. Most notably is the confounded variable task which seemed simply too difficult for most participants. About 19% of the participants appeared to engage in ANCOVA-like reasoning. A few informal protocols performed at the end of the testing battery were particularly informative. Some participants appeared to have difficulty with "reading" the data table provided in the confounded variable task. Thus some participants may not have provided the correct answer due to issues not of

---

13 That thinking dispositions may be moderated by the level of cognitive capacity is an important issue not dealt with in this study. Faced with some complex logical or covariation problem, for example, dispositions to be impulsive or high in need for cognitive closure may not be detrimental to the cognitive system that lacks the capacity to implement the necessary algorithms or mental operations to begin with. Such a system has the option of either firing off a quick response or investing time and mental work in a more deliberative fashion. In either case the normative response is unlikely to be generated, but the latter option carries the additional cost of time and mental effort.
interest to the task. The data table could potentially have been laid out in a more understandable fashion. Other tasks may have benefited through the inclusion of more items (see above point)—although without dropping some tasks from the battery, this would have resulted in a testing session that is too long for participants. Furthermore, several of the tasks may be criticized as being too artificial and not relevant to everyday reasoning and judgment. The present research may have benefited from the inclusion of at least one more task resembling to a greater degree types of reasoning and judgments that occur in the real world. An example of a possible prior belief task may have been a simulated juror reasoning task.

Conclusions

As previously outlined, an important conjecture in the critical thinking research tradition is that prior belief should not dominate reasoning. A conflicting view is advocated by the contextualist research tradition. Arguing that cognition is very much context specific, the contextualist tradition views cognition as utilizing situation-specific strategies. Any separation of context-relevant beliefs from the reasoning context is seen to be maladaptive in the contextualist research tradition. Thus, participants with higher levels of cognitive ability should take advantage of the contextually relevant cues—such as prior beliefs. The present study demonstrated that the contextualist's claim that incorporating situation specific knowledge in reasoning is indeed important for effectiveness. The tendency to overuse the gender cue in the ecological height judgment task, i.e., belief biased judgment, was associated with both higher levels of cognitive ability and, importantly, more accurate tracking of the actual stimuli. An analogous result was also evident in the opinion prediction task. The projection of consensus was positively correlated with cognitive ability, and importantly, facilitated predictive accuracy. That this is true is accompanied by
an important caveat drawn from the critical thinking research tradition: When it would be inefficacious to project some context-relevant information, e.g., prior belief about a gender/height relationship, it is those participants with higher levels of cognitive ability that are more successful at avoiding this contextualized cue. Belief biased judgments in the matched height judgment task was associated with significantly lower levels of cognitive ability. Thus the flexibility to appropriately adjust the weight given to the very same context-relevant belief as a function of its diagnosticity was positively associated with general cognitive ability. A result also consistent with this point was evident in the social perception task of knowledge prediction. When there appeared to be cues of higher diagnosticity than that afforded by consideration of the own knowledge cue alone (a situation that was not the case in the opinion prediction task), the correlation between projection and cognitive ability disappeared— but the correlation between predictive accuracy and cognitive ability, nevertheless, remained significant.

As was true with the matched height judgment task, belief bias on all the remaining reasoning tasks—in which prior belief was pitted against accurate response—was negatively correlated with cognitive ability. This result is contrary to an extreme contextualist position which would view the tendency to avoid contextual information, including relevant prior beliefs, as maladaptive and hence unlikely to be the case for those with greater cognitive resources. A consistent finding across the several reasoning tasks employed in this research demonstrated that participants with higher levels of cognitive ability were less likely to project context-relevant beliefs when it would be inefficacious to do so.

That belief bias displayed some domain generality in the present research is also consistent with the notion of general critical thinking skills and incongruent with an extreme contextualist hypothesis of domain specificity. Belief bias was
not found to be specific to a domain, rather the results demonstrated that the
tendency to project the nondiagnostic belief cue in a deductive logic task was
significantly correlated with the tendency to project a nondiagnostic belief in a
perceptual task. Furthermore, the ability to reason effectively independently of
belief on tasks requiring the evaluation of arguments and covariation data was
also intercorrelated with these indices. Thus the ability to decouple contextual
information, such as prior beliefs, has at least some generality.

The results of the thinking disposition scales also has relevance for issues
discussed in the critical thinking literature. The actively open-minded thinking
(AOT) composite scale enjoyed some measure of success in this study. The AOT
composite continued to account for additional variance in both the AET and the
syllogisms test even after cognitive ability was statistically controlled. This in
itself is an interesting result. In terms of a tool of measurement, the thinking
disposition scales are quite distinct from measures of cognitive ability. They tap
underlying epistemic goals and values, such as epistemic absolutism,
identification of beliefs with the concept of self, willingness to be simply
dissipative of the opinions of others or evidence conflicting with held beliefs etc.

The fact that some of this variation in prior belief task performance is also
accounted for by cognitive ability does not take away the charge that actively open-
 minded dispositions--as advocated by critical thinking theorists--are conducive to
cognitive effectiveness. On the contrary, the existing correlation between
cognitive capacity and actively open-minded thinking dispositions suggests, at the
very least, that people with increased levels of cognitive ability recognize the
utility of these dispositions in their thinking and reasoning activities. Actively
open-minded dispositions were clearly linked to superior performance on some of
the prior belief tasks. Furthermore, that thinking dispositions did account for
additional variance on the AET and syllogisms test after cognitive ability is
statistically controlled suggests that these dispositions are not only the consequence of cognitive ability, but rather are determinants of effective reasoning which lie at a different level of cognitive analysis than cognitive ability.
References


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: Lawrence Erlbaum Associates.


Byrnes, J. P. (1995). Domain-specificity and the logic of using general ability as an


Campbell, J. D. (1986). Similarity and uniqueness: The effects of attribute type,
relevance, and individual differences in self-esteem and depression. *Journal

measures: A theoretical account of the processing in the Raven Progressive

studies.* Cambridge: Cambridge University Press.

*Intelligence, 24,* 25-52.

Caryl, P. G. (1994). Early event-related potentials correlate with inspection time
and intelligence. *Intelligence, 18,* 15-46.

Casscells, W., Schoenberger, A., & Graboys, T. (1978). Interpretation by
physicians of clinical laboratory results. *New England Journal of Medicine,
299,* 999-1001.

Review, 13,* 403-435.


Maimon, D. Nodine, & O'Connor (Eds.), *Thinking, reasoning, and writing* (pp.
147-159). New York: Longman.

psychodiagnostic observations. *Journal of Abnormal Psychology, 72,* 193-204.


Belief Bias


Belief Bias

JSAS Catalog of Selected Documents in Psychology, 7(100 (Ms. No. 1577)).
Belief Bias

Press.


Belief Bias


Norwood, NJ: Ablex.


Appendix A

**Syllogism Examples**

**Inconsistent Syllogisms (N = 8)**

1. Premise 1: All living things need water.
   Premise 2: *Roses need water*.
   Conclusion: Roses are living things. (invalid)

2. Premise 1: All things that are smoked are good for the health.
   Premise 2: Cigarettes are smoked.
   Conclusion: Cigarettes are good for the health. (valid)

**Consistent Syllogisms (N = 8)**

1. Premise 1: All things made of wood can be used as fuel.
   Premise 1: Gasoline is not made of wood.
   Conclusion: Gasoline cannot be used as fuel. (invalid)

2. Premise 1: All fish can swim.
   Premise 2: *Tuna are fish*.
   Conclusion: Tuna can swim. (valid)

**Neutral Syllogisms (N = 8)**

1. Premise 1: All opprobines run on electricity.
   Premise 2: *Jamtops run on electricity*.
   Conclusion: Jamtops are opprobines. (invalid)

2. Premise 1: All ramadions taste delicious.
   Premise 2: *Gumthorps are ramadions*.
   Conclusion: Gumthorps taste delicious. (valid)

*Note.* For each of the three type of syllogisms, four were logically valid and four were logically invalid.
Appendix B

Examples of Stimuli Used in Height Judgment Tasks
Appendix C

Examples of AET Items

Part I: Prior Belief Instrument

Students should have a stronger voice than the general public in setting university policies.

A=Strongly Disagree    B=Disagree    C=Agree    D=Strongly Agree

Part II: Argument Evaluation Instrument

Dale's belief: Students, not the general public, should have the ultimate voice in setting university policies.

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

Critic's counter-argument: Tuition covers less than one half the cost of an education at most universities (assume statement factually correct).

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 make the ultimate decisions.

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

A=Very Weak    B=Weak    C=Strong    D=Very Strong

Median experts’ evaluation (Argument Quality) = 1.00
Mean item agreement across participants (Prior Belief) = 3.16
Mean rebuttal evaluation across participants (Argument Evaluation) = 2.85
Appendix C (continued)

Part I: Prior Belief Instrument

Judges should sentence more juveniles to prison for their crimes.

\[ A=\text{Strongly Disagree} \quad B=\text{Disagree} \quad C=\text{Agree} \quad D=\text{Strongly Agree} \]

Part II: Argument Evaluation Instrument

Dale's belief: Judges should sentence more juveniles to prison for their crimes.

Dale's premise or justification for belief: More juvenile delinquents should be sentenced to prison because severe punishment will act as a deterrent to future criminal activities.

Critic's counter-argument: Youths who are sent to prison for relatively minor infractions have a greater likelihood of becoming adult criminals than youths who are placed on parole in a community rehabilitation program (assume statement factually correct).

Dale's rebuttal to Critic's counter-argument: Youths who commit even relatively minor infractions are more likely to engage in criminal activities as adults than nondelinquent youths (assume statement factually correct).

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

\[ A=\text{Very Weak} \quad B=\text{Weak} \quad C=\text{Strong} \quad D=\text{Very Strong} \]

Median experts' evaluation (Argument Quality) = 1.00
Mean item agreement across participants (Prior Belief) = 2.73
Mean rebuttal evaluation across participants (Argument Evaluation) = 2.24
Appendix D

Examples of Covariation Judgment Items

Part I: Prior Belief Instrument

1. A student believes that high self esteem is associated with leadership qualities. To what extent do you agree or disagree with this student’s belief? Circle a number from minus 10 to plus 10 on the following scale:

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10

strongly disagree neutral strongly agree

Part II: Covariation Judgment Instrument

Imagine that a researcher sampled 450 individuals and found that:

200 people with low self esteem were low in leadership qualities
100 people with low self esteem were high in leadership qualities
50 people with high self esteem were low in leadership qualities
100 people with high self esteem were high in leadership qualities

Please judge the nature and extent of the relationship between self esteem and leadership qualities in these data.

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10

strong negative association no association strong positive association

Actual Δp = .333
Appendix D (continued)

Part I: Prior Belief Instrument

3. A student believes that having siblings is associated with sociability. To what extent do you agree or disagree with this student’s belief?

-10  -9  -8  -7  -6  -5  -4  -3  -2  -1  0  +1  +2  +3  +4  +5  +6  +7  +8  +9  +10

strongly disagree  neutral  strongly agree

Part II: Covariation Judgment Instrument

Imagine that a researcher sampled 675 children and found that:

150 children had siblings and were not sociable B
150 children had siblings and were sociable A
75 children did not have siblings and were not sociable D
300 children did not have siblings and were sociable C

Please judge the nature and extent of the relationship between having siblings and sociability in these data.

-10  -9  -8  -7  -6  -5  -4  -3  -2  -1  0  +1  +2  +3  +4  +5  +6  +7  +8  +9  +10

strong negative association  no association  strong positive association

Actual Δp = -.300
Confounded Variable Task

(Note that both possible data sets are included here)

Fire departments in the city of Chicago have currently been embroiled in a contentious issue. The issue revolves around revising the standard selection tests used to screen prospective firefighters. These tests are designed to predict which job applicants would go on to become successful firefighters and which applicants would not. A recent suggestion has been made that in addition to the usual battery of tests, the Risky-Conservative Choice test (RCC test) also be administered. The RCC test has been shown to reliably determine an applicant’s disposition to take risks (risk-taker) or disposition not to engage in risk taking behaviour (risk-avoiding). It has been persuasively argued by many that risk taking firefighters are simply more willing to risk their own personal safety in effort to rescue others and thus make for more successful firefighters than risk avoiders. Two recent cases have been cited continuously by proponents of the RCC test as to why its use is necessary:

The first case involves a firefighter, John P., who was described by his fellow workers as a risk taker. John is also very much considered to be an outstanding firefighter by his superiors and colleagues. While fighting a fire several months ago, John was confronted with a dangerous "back draft" situation. Behind a closed bedroom door of a young unconscious boy raged a fire. Opening doors in such situations result in a tremendous thrust of fire in the immediate direction of the person opening the door. Firefighters are trained to deal with these situations, but there is absolutely no room for error. John didn't hesitate one moment in opening the door and was successful at rescuing the unconscious boy.

The above case is a stark contrast to an incident which occurred a couple weeks later. William F., a firefighter at another station in Chicago was considered a conservative guy who liked to play it safe. One night while fighting a fire, William was confronted with the exact same situation John had faced earlier. William, however, hesitated several moments before opening the door. The young boy recovered from that room later died in the hospital. It was widely believed that the young boy would have survived had William not hesitated in dealing with the risky back draft situation.

Give this a moment's thought. Knowing what you know about the nature of fire fighting, do you believe, generally speaking, that risk takers or risk avoiders make for more successful firefighters?__________________________________________________________________________

Give one good reason for your belief on this issue (in point form):

Please indicate on the below scale your own personal opinion:

1. I strongly agree that risk-taking firefighters will be better performers.
2. I agree that risk-taking firefighters will be better performers.
3. I mildly agree that risk-taking firefighters will be better performers.
4. I mildly disagree that risk-taking firefighters will be better performers.
5. I disagree that risk-taking firefighters will be better performers.
6. I strongly disagree that risk-taking firefighters will be better performers.
Appendix E (continued)

First Possible Data Set:

Recall the firefighter controversy in Chicago you read about. The city of Chicago has employed the services of a hiring consultant. Although the Risky-Conservative test (RCC) had not been used to screen applicants, it had been, however, administered for some time now on new recruits in anticipation of it’s possible future use. In addition to the RCC test classification, the consultant also had information regarding the subsequent success or failure of each of these fire fighters. The success and failure data was derived from job evaluations conducted a year after the applicants were hired. Below you will find the actual data given to the hiring consultant. The data was collected from fire stations that serve either the difficult industrial fires (Industrial fire stations) or ordinary house fires (Residential fire stations).

The below table summarizes the data in the following manner: The rows correspond to the RCC test’s classification of whether the fire fighter had been assessed as a risk taker or as risk avoiding. The columns indicate information regarding the break down of successful versus failed fire fighters. You are first given the total overall success and failure information. Next is the success and failure information for residential fire stations only, followed by the success and failure information for industrial fire stations only. Note that the total overall information is simply the residential and industrial fire station information combined.

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Total</th>
<th>Residential Fire Stations</th>
<th>Industrial Fire Stations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall</td>
<td>Succ.</td>
<td>Fail</td>
</tr>
<tr>
<td>Risk avoiding</td>
<td>10</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Risk takers</td>
<td>15</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

On the basis of the above table data alone, please circle the item on the below scale which best describes the relationship between risk taking and firefighter success:

1. Risk taking firefighters are markedly superior to risk avoiding firefighters.
2. Risk taking firefighters are much better than risk avoiding firefighters.
3. Risk taking firefighters are better than risk avoiding firefighters.
4. Risk taking firefighters are slightly better than risk avoiding firefighters.
5. There is no difference between risk taking and risk avoiding firefighters.
6. Risk avoiding firefighters are slightly better than risk taking firefighters.
7. Risk avoiding firefighters are better than risk taking firefighters.
8. Risk avoiding firefighters are much better than risk taking firefighters.
9. Risk avoiding firefighters are markedly superior to risk taking firefighters.
Appendix E (continued)

Second Possible Data Set:

Recall the firefighter controversy in Chicago you read about. The city of Chicago has employed the services of a hiring consultant. Although the Risky-Conservative test (RCC) had not been used to screen applicants, it had been, however, administered for some time now on new recruits in anticipation of its possible future use. In addition to the RCC test classification, the consultant also had information regarding the subsequent success or failure of each of these fire fighters. The success and failure data was derived from job evaluations conducted a year after the applicants were hired. Below you will find the actual data given to the hiring consultant. The data was collected from fire stations that serve either the difficult industrial fires (Industrial fire stations) or ordinary house fires (Residential fire stations).

The below table summarizes the data in the following manner: The rows correspond to the RCC test's classification of whether the fire fighter had been assessed as a risk taker or as risk avoiding. The columns indicate information regarding the break down of successful versus failed fire fighters. You are first given the total overall success and failure information. Next is the success and failure information for residential fire stations only, followed by the success and failure information for industrial fire stations only. Note that the total overall information is simply the residential and industrial fire station information combined.

<table>
<thead>
<tr>
<th></th>
<th>Total overall</th>
<th>Residential fire stations</th>
<th>Industrial fire stations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Succ.</td>
<td>Fail</td>
<td>Succ.</td>
</tr>
<tr>
<td>Risk takers</td>
<td>10</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Risk avoiding</td>
<td>15</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

On the basis of the above table data alone, please circle the item on the below scale which best describes the relationship between risk taking and firefighter success:

1. Risk taking firefighters are **markedly superior** to risk avoiding firefighters.
2. Risk taking firefighters are **much better** than risk avoiding firefighters.
3. Risk taking firefighters are **better** than risk avoiding firefighters.
4. Risk taking firefighters are **slightly better** than risk avoiding firefighters.
5. There is **no difference** between risk taking and risk avoiding firefighters.
6. Risk avoiding firefighters are **slightly better** than risk taking firefighters.
7. Risk avoiding firefighters are **better** than risk taking firefighters.
8. Risk avoiding firefighters are **much better** than risk taking firefighters.
9. Risk avoiding firefighters are **markedly superior** to risk taking firefighters.
Appendix F

Vocabulary Checklist

Below you will see a list of 60 letter strings. Some of the strings are actual words and some are not. You are to read through the list of items and indicate whether or not you think the letter string is a word by putting a check mark next to those that you know to be words. Do not guess, but only check those who you know to be words.

1. absolution ____ 31. neotatin ____
2. arrate ____ 32. niche ____
3. asinine ____ 33. nonquasity ____
4. audible ____ 34. nuance ____
5. ceiloplaty ____ 35. nitrous ____
6. clandestine ____ 36. optimize ____
7. comectial ____ 37. plabage ____
8. concurrent ____ 38. polarity ____
9. confluence ____ 39. potomite ____
10. connote ____ 40. purview ____
11. denotation ____ 41. recidivism ____
12. denouement ____ 42. reportage ____
13. disconcert ____ 43. reverent ____
14. disler ____ 44. rochead ____
15. dropant ____ 45. seblemant ____
16. epicurean ____ 46. sheal ____
17. eventuate ____ 47. sparkhouse ____
18. fusigenic ____ 48. stratagem ____
19. gustation ____ 49. subjugate ____
20. heuristic ____ 50. substratum ____
21. hyplexion ____ 51. suffuse ____
22. ineffity ____ 52. tenacious ____
23. inflect ____ 53. tradured ____
24. inundate ____ 54. tuncier ____
25. irksome ____ 55. ubiquitous ____
26. lacuna ____ 56. uction ____
27. langour ____ 57. unmanal ____
28. laudatory ____ 58. wanderlust ____
29. litany ____ 59. waterfowl ____
30. metenetion ____ 60. xenophobia ____
Appendix G

Print Exposure Measures

1. **Author Recognition Test (ART):**

Below you will see a list of names. Some of the people in the list are popular writers (of books, magazine articles, and/or newspaper columns) and some are not. You are to read the names and put a check mark next to the names of those individuals who you know to be writers. Do not guess, but only check those who you know to be writers. Remember, some of the names are people who are not popular writers, so guessing can easily be detected.

<table>
<thead>
<tr>
<th>1. Lauren Adamson</th>
<th>37. Robert Fulghum</th>
<th>73. K. Warner Schaie</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Eric Amsel</td>
<td>38. Howard Gardner</td>
<td>74. Sidney Sheldon</td>
</tr>
<tr>
<td>4. Isaac Asimov</td>
<td>40. Andrew Greeley</td>
<td>76. Danielle Steel</td>
</tr>
<tr>
<td>5. Margaret Atwood</td>
<td>41. Frank Gresham</td>
<td>77. Mark Strauss</td>
</tr>
<tr>
<td>6. Margarita Azmitia</td>
<td>42. John Grisham</td>
<td>78. Amy Tan</td>
</tr>
<tr>
<td>7. Oscar Barbarin</td>
<td>43. Alex Haley</td>
<td>79. Alvin Toffler</td>
</tr>
<tr>
<td>8. Reuben Baron</td>
<td>44. Frank Herbert</td>
<td>80. J. R. R. Tolkien</td>
</tr>
<tr>
<td>13. Elliot Blass</td>
<td>49. Frank Keil</td>
<td>85. Steve Yussen</td>
</tr>
<tr>
<td>15. Dale Blyth</td>
<td>51. Dean Koontz</td>
<td></td>
</tr>
<tr>
<td>16. Hilda Borko</td>
<td>52. Judith Krantz</td>
<td></td>
</tr>
<tr>
<td>17. Barbara Cartland</td>
<td>53. Louis L'Amour</td>
<td></td>
</tr>
<tr>
<td>18. Agatha Christie</td>
<td>54. Margaret Laurence</td>
<td></td>
</tr>
<tr>
<td>19. Tom Clancy</td>
<td>55. Reed Larson</td>
<td></td>
</tr>
<tr>
<td>21. James Clavell</td>
<td>57. C. S. Lewis</td>
<td></td>
</tr>
<tr>
<td>22. Jackie Collins</td>
<td>58. Lynn Liben</td>
<td></td>
</tr>
<tr>
<td>24. Stephen Coonts</td>
<td>60. Hugh Lytton</td>
<td></td>
</tr>
<tr>
<td>25. Edward Cornell</td>
<td>61. Franklin Manis</td>
<td></td>
</tr>
<tr>
<td>27. Diane Cuneo</td>
<td>63. James Michener</td>
<td></td>
</tr>
<tr>
<td>28. Denise Daniels</td>
<td>64. James Morgan</td>
<td></td>
</tr>
<tr>
<td>29. Robertson Davies</td>
<td>65. Alice Munro</td>
<td></td>
</tr>
<tr>
<td>30. Geraldine Dawson</td>
<td>66. Scott Paris</td>
<td></td>
</tr>
<tr>
<td>32. W. Patrick Dickson</td>
<td>68. M. Scott Peck</td>
<td></td>
</tr>
<tr>
<td>33. Robert Emery</td>
<td>69. David Perry</td>
<td></td>
</tr>
<tr>
<td>34. Frances Fincham</td>
<td>70. Anne Rice</td>
<td></td>
</tr>
<tr>
<td>35. Timothy Findley</td>
<td>71. Mordecai Richler</td>
<td></td>
</tr>
<tr>
<td>36. Martin Ford</td>
<td></td>
<td>72. Miriam Sexton</td>
</tr>
</tbody>
</table>
Appendix G (continued)

Print Exposure Measures

2. Magazine Recognition Test (MRT):

Below you will see a list of titles. Some of them are the names of actual magazines and some are not. You are to read the names and put a check mark next to the names of those that you know to be magazines. Do not guess, but only check those that you know to be actual magazines. Remember, some of the titles are not those of popular magazines, so guessing can easily be detected.

1. Allure
2. Architecture Today
3. Atlantic
4. Better Homes & Gardens
5. Business Week
6. Byte
7. Car & Driver
8. Chatelaine
9. Consumer Reports
10. Cosmopolitan
11. Digital Sound
12. Ebony
13. Electrical & Mechanical News
14. Elle
15. Elliot
16. Esquire
17. Essence
18. Family Circle
19. Field & Stream
20. Flare
21. Fitness Today
22. Forbes
23. Galactic Digest
24. Gentleman's Quarterly
25. Girl Weekly
26. Glamour
27. Good Housekeeping
28. Guitar World
29. Harper's Magazine
30. Health & Life
31. Home Finance
32. Hot Rod
33. House & Garden
34. Hunters
35. Illustrated Science
36. Inside Sports
37. Interview
38. Jet
39. Ladies Home Journal
40. Life
41. Maclean's
42. Madame
43. Mademoiselle
44. Market Trends
45. McCall's Magazine
46. Modern Family
47. Motor Sports
48. Motor Trend
49. Mountain and Stream
50. Music Weekly
51. National Geographic
52. Neuberger Review
53. New Yorker
54. Newsweek
55. Omni
56. Outdoor Life
57. Outdoor Times
58. Pacific World
59. People
60. Personal Psychology
61. Popular History
62. Popular Science
63. Premiere
64. Psychology Today
65. Public Policy Review
66. Reader's Choice
67. Reader's Digest
68. Recreation Today
69. Redbook
70. Road & Track
71. Rolling Stone
72. Safeco News Service
73. Sassy
74. Saturday Night
75. Science Quest
76. Science Reader
77. Scientific American
78. Self
79. Seventeen
80. Software Development
81. Spin
82. Sports Illustrated
83. Stock and Bond Digest
84. Technology Digest
85. This Magazine
86. Thrasher
87. Time
88. Tools and Repair
89. Town & Country
90. Urban Scene
91. Vanity Fair
92. Vogue
93. Vox
94. Wellesley