Individual Differences in Thinking, Reasoning, and Decision Making

Keith E. Stanovich, Walter C. Sá, and Richard F. West

The literature on individual differences in thinking and reasoning is quite scattered and does not form a coherent research paradigm. Instead, it rides parasitically alongside other research programs and other theoretical controversies such as the mental models versus mental-logic debate. Despite its nonprogrammatic nature, work on individual differences in reasoning has begun to provide some insights that would not have been available had the field ignored this important conceptual and methodological tool. In this chapter, we wish to highlight some examples of how work on individual differences serves to clarify issues in the cognitive psychology of reasoning. We will begin by discussing some examples that arise from within some of the well-developed areas in the reasoning literature including: mental logic, mental models, componential analysis, and crosscultural studies in reasoning. We will then discuss three programmatic themes from our own research studies on individual differences to show how we have attempted to demonstrate how individual differences research can lead to cumulative progress on theoretical issues of general importance.

MENTAL LOGIC AND INDIVIDUAL DIFFERENCES

Perhaps the most notable investigation of individual differences carried out within the mental logic (Braine & O’Brien, 1998; Rips, 1994), or mental rules, framework is that of Rips and Conrad (1983). They examined individual differences in reasoning with the propositional connectives if, and, or, and neither. Their work introduced the useful strategy of distinguishing individual differences in reasoning that result from random noise (e.g., temporary distractions and attentional lapses), from those differences deriving from differences in real deductive reasoning competence (see...
Individual Differences in Reasoning

Stanovich, 1999, and Stein, 1996, for discussions of the use of the competence/performance distinction in reasoning). Rips and Conrad (1983) argued that reasoning performance differences that result from the latter (but not the former) should demonstrate an appreciable degree of stability. This stability in performance was observed in the Rips and Conrad study. Participants’ performance patterns demonstrated stability across various reasoning tasks, and across a six-week period of time—precisely the pattern expected of individual differences grounded in deductive reasoning competence.

Having demonstrated that individual differences in deductive paradigms were not merely due to idiosyncratic factors, Rips and Conrad next explored the factors that led to systematic variation in performance. Specifically, they examined argument-by-subject interactions—which arguments sharing a particular logical property were consistently judged to be valid by one group of subjects and consistently judged as invalid by yet another group of subjects. For example, relatively consistently one group of subjects handled disjunctive introduction in the exclusive sense (p or q is true in the case where exactly one of the disjuncts is true), whereas another group of subjects consistently handled disjunctive introduction in the inclusive sense (p or q is true in the case where either one or both of the disjuncts is true).

From their investigation, Rips and Conrad (1983) concluded that a view of mental logic as a universal set of rules is inaccurate. Although some universal rules may exist, other rules may not be universal features of human cognitive architecture. In addition, Rips and Conrad argued that deficiencies in mental logic may be amendable through education. They found that subjects who had had training in logic were more likely to treat several important logical properties (including disjunctive introduction) in a manner congruent with the propositional calculus. As mental-logic theorists however, Rips and Conrad (1983) do maintain that despite the individual differences obtained in the instantiation of various logical rules all subjects carry out the processes involved in deduction by using some form of mental logic (i.e., rules).

MENTAL MODELS AND INDIVIDUAL DIFFERENCES

Mental-model accounts of reasoning (Johnson-Laird, 1999; Johnson-Laird & Byrne, 1991, 1993) offer yet a different mechanism underlying the processes involved in deductive reasoning. The mental-rules account views mental processes as operating directly on the form of the premises. Mental-models accounts, in contrast, emphasize the construction of intermediate representations between the given premises and a formulation of a conclusion. Under the mental-models view, an individual transforms the premises of an argument into a representation of individual tokens or a set of relationships that is consistent with those premises—that is, a mental model. This first representation is used in formulating a tentative conclusion. Additional mental models that are consistent with the premises may or may not be formed subsequent to this initial model. The construction of additional models (where possible) is desirable however because they will often demonstrate that the tentative conclusion is in fact invalid.

Early on in the development of this account, proponents of mental models acknowledged the existence of individual differences in reasoning (Johnson-Laird, 1983, Johnson-Laird & Bara, 1984). Writing in reference to syllogistic reasoning, Johnson-Laird and Bara (1984) noted that “there are undoubtedly differences from one individual to another in the way in which their participants made syllogistic inferences. . . . Certain, the cause of individual differences is a major problem that remains to be solved” (p. 50). One source of individual differences often implicated within the mental-model account is variation in working memory capacity. For example, Johnson-Laird and Byrne (1992) have pointed out that the model theory can “account for these individual differences in terms of such factors as the processing capacity of working memory” (p. 62). As Johnson-Laird and colleagues (Johnson-Laird, 1983, 1999; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991, 1993; see also Evans, Newstead, & Byrne, 1993) have emphasized, individual differences under the mental-models account may arise because some individuals more completely flesh out their models and because some individuals are able to keep multiple models in mind better than other individuals.

In short, because many problem-solving situations optimally can be evaluated by constructing all of the mental models that are consistent with the premises as presented, and because constructing all of the models might be limited by working memory constraints, one would expect the latter to predict performance on reasoning tasks. As expected under a mental-models account, performance on syllogistic reasoning tasks deteriorates in proportion to the number of alternative models that are required to solve the problem (Johnson-Laird & Steedman, 1978; see also Johnson-Laird, 1983). Furthermore, significant correlations have been obtained between measures of working memory capacity and performance in solving syllogisms (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984; Bara, Bucciarelli, & Johnson-Laird, 1995; Bucciarelli & Johnson-Laird, 1999).

Bara et al. (1995) studied syllogistic inference in children, adolescents, and adults and identified five basic cognitive factors that affect performance: the interpretation of quantifiers; the referential integration of assertions across the premises; the search for counterexamples; the ability to notice identities between constructed models; and the processing capacity of working memory. Each of these cognitive factors was separately operationalized by performance on a task designed to tap that particular
component. For example, an operationalization of the fifth factor (i.e., the processing capacity of working memory) was attained by using the common methodology of having participants recall a series of spoken digits—both forward and backward trials were used. The criterion variable in the Bara et al. (1995) investigation was measured by having participants generate conclusions for several syllogisms. A series of regression analyses revealed the ability to notice identities between constructed models and the processing capacity of working memory accounted for significant variation in syllogistic inference. While working memory capacity and the perception of identities account for a substantial portion of the variance in syllogistic reasoning performance (39% in the Bara et al., 1995, study), Bucciarelli and Johnson-Laird (1999) caution that “we are far from answering this important question” (p. 300).

QUALITATIVE INDIVIDUAL DIFFERENCES IN STRATEGIES: MENTAL MODELS VERSUS MENTAL RULES

A common research strategy used by both advocates of mental models and advocates of mental-logic approaches is to pit the predictions of these two opposing views against one another. Advocates of each have sought a universal reasoning mechanism for deductive reasoning. The fact that both accounts have received some support suggests that each view may be correct under certain types of conditions or for certain individuals. The reasoning of different individuals might be best characterized by a mental model or a mental-logic theory.

Such a conclusion is suggested by an important study conducted by Galotti, Baron, and Sabini (1986). In line with the popular strategy, Galotti et al. (1986) pitted the predictions of a mental-models account against those of a mental-rules model of syllogistic reasoning performance. Their first experiment examined whether the differences between poor and good reasoners occur early or late in their reasoning process. The two theoretical views hold different expectations in this regard. The mental-model account views individuals as constructing an initial mental model that is consistent with the premises. Errors in reasoning occur because individuals fail to construe the requisite number of additional premise-consistent models that are required to evaluate the validity of a conclusion. Thus poor and good reasoners can be seen as differing in the latter stages of the reasoning process. In contrast, the mental-logic account views reasoners as applying a fixed repertoire of rules directly to the premises. This scenario is consistent with the notion that errors occur early in the reasoning process.

Galotti et al. (1986) had their subjects provide “could be true” conclusions to syllogistic arguments on some trials and “must be true” conclusions on other trials. Since mental-model accounts view both good and poor reasoners as first engaging in the construction of a one premise-consistent model (i.e., a could be true conclusion), good and poor reasoners should not differ in their generation of a could be true conclusion that turns out to be an error under the instruction of detecting necessarily true conclusions. On the other hand, the mental-models account predicts that subjects will differ when asked to generate must be true conclusions. Good reasoners will distinguish this instruction from could be true and accordingly generate more premise-consistent models. The distinction becomes particularly pertinent for “nothing follows” (NF) syllogisms because they require the construction of two or more mental models. Thus, on the mental-models account, the errors and latencies of good reasoners should resemble poor reasoners in the could be true trials, and differ from poor reasoners in the must be true trials. Although rules accounts make no specific predictions about error or latencies, an observed lack of differences—clearly inconsistent with the mental-models account—can be viewed as consistent with the rules account.

The results of the Galotti et al. (1986) experiment revealed that good and poor reasoners did not actually differ in accuracy when generating the could be true conclusions. This finding is consistent with the mental-models account. More importantly, the good and poor reasoners did differ on the must be true conclusions that used NF syllogisms. This finding was also consistent with the mental-models account. Although the data pattern involving group accuracy was consistent with the mental-models account, the latency data were not completely consistent with its predictions, because the good reasoners took reliably more time than the poor reasoners only under some of the must be true conditions.

Other components of the Galotti et al. (1986) investigation showed that the thinking of at least some subjects was consistent with a mental-logic account. Galotti et al. (1986) recalled most of their subjects to another testing session where they were asked to think aloud as they worked through syllogisms. The protocols revealed that some subjects used deduction rules in solving the syllogisms—an observation obviously consistent with a mental-logic account. Good reasoners were found to be more likely to announce a rule than the poor reasoners. In addition, poor reasoners were more likely to misinterpret “some” to mean “not all.” This latter finding is in line with Rips and Conrad's (1985) demonstration that stable individual differences can be found in how subjects interpret logical connectives. In Experiment 2 of their investigation, Galotti et al. (1986) added an expert group in addition to the groups of poor and good reasoners. The expert group consisted of graduate students in psychology and computer science who had had some experience studying logic. The performance of the experts was best accommodated by a mental-rules account.

Many aspects of the Galotti et al. (1986) investigation seem to suggest that neither a mental-rules account nor a mental-models account merits the status of being deemed the fundamental reasoning mechanism. Instead,
Galotti et al. (1986) conclude their paper by suggesting that more research needs to be done to determine what circumstances lead people to apply rules or construct mental models. Even more importantly, what their investigations suggest is that individual differences might arise from reasoning processes that are qualitatively different from each other. Individual differences might encompass more than just differing parameter settings but instead might reflect differences in the basic cognitive mechanisms applied to the problem.

INDIVIDUAL DIFFERENCES AND COMPONENTIAL ANALYSIS

A similar conclusion can be derived from the earlier work by Sternberg (1977, 1980; Schustack & Sternberg, 1981; Sternberg & Weil, 1980). In a componential analysis of analogical reasoning, Sternberg (1977) identified six separate cognitive sources of individual differences. The sixth category is particularly relevant here since it implicates differences in the mental representation on which components act—that is, it implicates qualitative differences. Sternberg found that some of his subjects represented information primarily in a linguistic mode, whereas others represented the same information primarily in a spatial mode. Similarly, in further investigations involving linear syllogisms, some subjects were found to operate on array-like representations (more like mental models), others seemed to be operating upon linguistic-like representations that were more like mental rules. Some used a mix of the two mechanisms.

Outcomes like those of Galotti et al. (1986) and Sternberg’s work using componential analysis to identify individual differences in reasoning tasks provides the context for a critical review by Roberts (1993) of the search for basic reasoning mechanisms. In his analysis, Roberts (1993) highlights the importance of individual differences and the futility of searching for reasoning patterns that are universal. Disputes such as those between the mental models and mental rules view are unlikely to result in an unequivocal winner of the theoretical contest. Instead, an outcome like that of Sternberg’s componential analyses seems more likely—subgroups of individuals will be identified that among themselves show some common reasoning strategies. In addition, the choice of strategy can be viewed as partially a function of the task selected by the experimenter. Thus for example, tasks that would exceed working-memory constraints under a mental-model strategy may be conducive to the application of rules. The facilitation of mental-models use in syllogistic reasoning can be achieved by presenting the premises one at a time; ideally in a verbal mode as opposed to visual presentation (see Roberts, 1993). Conversely, the presentation of the problem premises simultaneously in written form encourages deductive rule strategies.

INDIVIDUAL DIFFERENCES IN REASONING EXPLORED FROM A CROSSCULTURAL PERSPECTIVE

The final tradition which provides a perspective for studies of individual differences in reasoning is one in which established reasoning tasks and paradigms are explored from a crosscultural perspective. The most notable recent programmatic exploration of individual differences has been conducted by Richard Nisbett and colleagues (e.g., Nisbett, Peng, Choi, & Norenzayan, 2001; Norenzayan & Nisbett, 2000; Peng & Nisbett, 1999). They have attempted to empirically demonstrate the existence of different reasoning styles in Eastern and Western cultures. Although controversial (Hong, Morris, Chiu, & Benet-Martinez, 2000), the work of Nisbett and colleagues seems to have at least suggested that individuals grounded in Western cultures seem to prefer analytic thinking styles. These styles are characterized by a focus on object attributes, categorization based on these attributes, the use of explicit rules and formal logic, and an intolerance for contradiction. In contrast, Nisbett’s work has suggested that individuals grounded in Eastern cultures prefer holistic styles characterized by a focus on context. In short, Western cultures can be seen as emphasizing decontextualized modes (see Stanovich, 1999), whereas Eastern cultures place importance on contextualized modes.

Nisbett and colleagues have examined these trends using a variety of tasks, but their work on reasoning about contradiction will serve as an illustration. Peng and Nisbett (1999) found that a group of Chinese subjects preferred dialectical proverbs containing apparent contradictions (e.g., “too humble is half proud”) more than their American counterparts. This finding was replicated in an additional experiment that used proverbs foreign to both the Chinese and Americans. Consistent differences between their Chinese and American subjects were also found in a third experiment on resolving social contradictions. Chinese and American subjects were provided with scenarios that outlined a conflict between two parties and were asked to resolve this conflict. Responses were classified as either dialectical or nondialectical strategies. Dialectical resolutions were those that addressed the issues from both sides of the dispute. Nondialectical resolutions typically assigned fault to one of the parties. The results showed the majority of their American subjects engaged in a nondialectical resolution, whereas Chinese subjects were more likely to assign some level of blame or fault to both parties and to resolve the conflict by some form of compromise.

Peng and Nisbett (1999) demonstrated differences in preferences for formal argumentation in a fourth experiment. Both logical arguments and dialectical arguments were presented to their subjects. The logical argument applied the rule of noncontradiction (no statement can be both true and false at the same time), whereas the dialectical argument applied the
principle of holism (emphasis on the idea that nothing is isolated and independent, instead, everything is connected). The American subjects were found to prefer arguments that contained the law of noncontradiction, whereas the Chinese subjects showed the opposite preference. Finally, in a fifth experiment, Peng and Nisbett (1999) demonstrated an analogous set of cultural differences when reasoning about scientific information and conclusions. In this and other studies (Nisbett et al., 2001; Norenzayan & Nisbett, 2000) cultural differences are associated with individual differences in the reasoning strategies and styles used to engage with a variety of reasoning and problem-solving tasks.

SYSTEMATIC INVESTIGATION OF THREE ISSUES IN REASONING WITH AN INDIVIDUAL DIFFERENCES APPROACH

As the previous sampling of research indicates, the importance of individual differences across a variety of reasoning paradigms has not gone unnoticed. However, with the exception of the research of the Nisbett group, work on individual differences in reasoning has lacked thematic focus. In the remainder of this chapter we will try to illustrate how such focus may be achieved by illustrating how we have used inferences from patterns of individual differences to address three conceptual issues in the psychology of reasoning and decision making.

The first issue we have addressed is the issue of whether, across the many tasks in the reasoning literature, there are commonalities in the performance patterns. That is, is there such a thing as a general tendency to reason optimally – or does the tendency toward normatively appropriate responding display extreme domain specificity (the reigning assumption in cognitive science and developmental psychology throughout the 1990s; see Ceci, 1996; Hirschfeld & Gelman, 1994; Samuels, 1996)? Secondly, what are the roles of cognitive ability and thinking dispositions in accounting for variance in reasoning ability? Finally, an innovative but controversial aspect of our research program is the attempt to use individual differences to adjudicate disputes about the appropriate normative models to apply to experimental tasks in the heuristics and biases literature.

ISSUE 1: IS THERE DOMAIN GENERALITY IN REASONING TENDENCIES ACROSS TASKS?

The starting point for our research on this question was the so-called heuristics and biases literature, now several decades in the making. Our point of demarcation was the fact that literally hundreds of empirical studies conducted over nearly three decades – have firmly established that people's responses often deviate from the performance considered normative on many reasoning tasks. For example, people display confirmation bias, they test hypotheses inefficiently, they do not properly calibrate degrees of belief, they overproject their own opinions onto others, they allow prior knowledge to become implicated in deductive reasoning, and they display numerous other information processing biases (for summaries of the large literature, see Baron, 1998, 2000; Dawes, 1998; Evans et al., 1993; Evans & Over, 1996; Hastie & Dawes, 2001; Johnson-Laird, 1999; Kahneman & Tversky, 2000; Manktelow, 1999; Medin & Bazerman, 1999; Nickerson, 1998; Shafir & Tversky, 1995; Stanovich, 1999). However, for a long time, one aspect of this empirical literature received little attention – that although the average person in these experiments might well display a reasoning error, some did not. We wondered whether, across various tasks, these might tend to be the same people, thus indicating some domain generality in thinking as operationalized by giving the standard normative response on reasoning tasks. It is precisely this question that we have addressed in a number of studies.

A typical set of results, taken from one of our studies (Stanovich & West, 1998), is displayed in Table 14.1. The tasks investigated here included a syllogistic reasoning task in which the believability of the conclusion contradicted logical validity (Evans, Barston, & Pollard, 1983). Next were five selection tasks (Wason, 1966). All employed nondeontic content, were difficult to solve, and thus were the type of problem about which there has been considerable debate in the research literature (Manktelow, 1999; Newstead & Evans, 1995). The third task was derived from the literature on statistical reasoning, and was inspired by the work of Nisbett and Ross (1980). Their work suggests the tendency for human judgment to be overly influenced by vivid but unrepresentative personal and case evidence and to be underinfluenced by more representative and diagnostic, but pallid, statistical evidence. The fourth task was an argument evaluation task (Stanovich & West, 1997) that taps reasoning skills of the type studied in the informal reasoning literature (see Baron, 1995; Klaczynski, 2000; Klaczynski, Gordon, & Fauth, 1997; Kuhn, 2001; Perkins, 1989). Importantly, to do well on it, one has to adhere to the Bayesian-like structure not to implicate prior belief in the evaluation of evidence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllogisms</td>
<td>.363**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection task</td>
<td></td>
<td>.258**</td>
<td></td>
</tr>
<tr>
<td>Statistical reasoning</td>
<td>-.334**</td>
<td>-.310**</td>
<td>.117</td>
</tr>
<tr>
<td>Argument Evaluation</td>
<td>-.340**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** = p < .001; all two-tailed; Ns = 188 to 195.
As Table 14.1 indicates, five of the six correlations among tasks were significant at the .001 level. The syllogistic reasoning task displayed significant correlations with each of the other three tasks, as did the selection task. The only correlation that did not attain significance was that between performance on the argument evaluation task and statistical reasoning. Although the highest correlation obtained was that between syllogistic reasoning and selection task performance (.363), correlations almost as strong were obtained between tasks deriving from the deductive reasoning literature (syllogistic reasoning, selection task) and inductive reasoning literature (statistical reasoning). These systematic tendencies may be even greater than indicated here because of the modest reliability of most of the tasks. The positive manifold displayed by the tasks suggests that some domain generality is present and that it makes at least some sense to talk about a general tendency toward rational thought.

Similar relationships were obtained in a second study (see Stanovich & West, 1998c, Study 2) where we expanded the selection of tasks by adding further reasoning tasks to those used in the original study. Added to our multivariate battery was a covariation detection task modeled on the work of Wasserman, Dorrer, and Kao (1990). Next was a hypothesis testing task modeled on Tschirgi (1980) in which the score on the task was the number of subjects attempted to test a hypothesis in a manner that confounded variables. Outcome bias was measured using tasks introduced by Baron and Hershey (1988). This bias is demonstrated when subjects rate a decision with a positive outcome as superior to a decision with a negative outcome even when the information available to the decision maker was the same in both cases. Finally, if only one bias refers to the tendency for people to have differential responses to outcomes based on the differences in counterfactual alternative outcomes that might have occurred (Epstein, Lipson, Holstein, & Huh, 1992; Miller, Turnbull, & McFarland, 1990). The bias is demonstrated when subjects rate a decision leading to a negative outcome as worse than a control condition when the former makes it easier to imagine a positive outcome occurring.

Modest but significant correlations were found among these new measures and between them and several of the tasks displayed in Table 14.1. Other work by our research group (Sá & Stanovich, 2001; Sá, West, & Stanovich, 1999; Stanovich & West, 1998a, 1998b, 1998c, 1998d, 1999, 2000b) and others (Parker & Fischhoff, 2000; Slugoski & Wilson, 1998) has converged in indicating some domain generality to normative responding on reasoning and decision-making tasks. Certainly there is considerable domain specificity involved here as well, but this was to be expected given the current emphasis on the domain specificity of cognitive functioning in cognitive science (Cosmides & Tooby, 1994; Hirschfeld & Gelman, 1994; Samuels, 1998; Sperber, 1994, 2000). Thus, the critical issue here is whether any domain generality at all can be detected, and our investigations have indicated that across a wide domain of tasks it does make sense to speak about a generalized tendency toward normative responding on reasoning tasks.

**ISSUE 2: COGNITIVE ABILITY AND THINKING DISPOSITIONS AS PREDICTORS OF INDIVIDUAL DIFFERENCES IN REASONING**

What is the source of the common variance in performance on these tasks? One obvious source of common variance is general cognitive ability. That is, the common variance we observed might simply be due to general computational limitations that the tasks imposed on limited-capacity cognitive structures that vary among individuals. In our studies, we have operationalized general cognitive capacity in terms of well-known cognitive ability and aptitude tasks such as the Scholastic Assessment Test (SAT), Raven Matrices, and various vocabulary and reading comprehension tests. All are known to load highly on psychometric g. Most of the analyses throughout this chapter will focus on SAT Total scores.

It should be emphasized that other measures of cognitive ability not involving verbal problem solving (such as the Raven matrices and checklist vocabulary measures) produced largely redundant findings. The correlations observed are not due to verbal reasoning items on the SAT bearing similarities to items on the thinking and reasoning tasks—an issue that we dealt with at length in the responses to the commentators on our “Behavioral and Brain Sciences” article on the rationality debate (Stanovich & West, 2000b). The reasoning and heuristics and biases literature from which we have drawn our tasks encompasses vastly more than the syllogistic reasoning literature—where, granted, items may resemble verbal reasoning on aptitude tests (although even here, the latter never contain a belief-bias component). The entire set of tasks used were varied and extended well beyond syllogistic reasoning. The choice between a vivid case and a statistical fact on an inductive reasoning problem (Nisbett & Ross, 1980; Stanovich & West, 1998c) is nothing like an item on the SAT; neither is an informal reasoning item in which prior belief must be ignored (Stanovich & West, 1997): neither is the combining of a diagnostic indicator and base-rate information (Stanovich & West, 1998d); neither is covariation assessment in the face of belief bias (Levin, Wasserman, & Kao, 1993; Stanovich & West, 1998d); and so on down a long list of tasks. The correlations between general ability and performance on such measures are of genuine interest.

The top half of Table 14.2 indicates the magnitude of the correlation between SAT total scores and the four reasoning tasks from Study 1 of Stanovich and West (1998c). SAT scores were significantly correlated with performance on all four rational thinking tasks. The correlation with syllogistic reasoning was the highest (.470) and the other three correlations were
TABLE 14.2. Correlations Between Performance on the Reasoning Tasks and SAT Total Score

<table>
<thead>
<tr>
<th>Data from Study 1 of Stanovich and West (1998c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syllogisms</strong></td>
</tr>
<tr>
<td><strong>Selection task</strong></td>
</tr>
<tr>
<td><strong>Statistical reasoning</strong></td>
</tr>
<tr>
<td><strong>Argument Evaluation</strong></td>
</tr>
</tbody>
</table>

**Replication and Extension (Study 2 of Stanovich & West, 1998c)**

| **Syllogisms**                                | .410* |
| **Statistical reasoning**                     | .376* |
| **Argument evaluation task**                  | .371* |
| **Covariation detection**                     | .239* |
| **Hypothesis testing bias**                   | .223* |
| **Outcome bias**                              | .172* |
| **If/Only thinking**                          | .208* |
| **Composite #1**                              | .530* |
| **Composite #2**                              | .383* |
| **Composite, All Tasks**                      | .547* |

*Note: Composite #1 = standard score composite of performance on argument evaluation task, syllogisms, and statistical reasoning. Composite #2 = standard score composite of performance on covariation judgment, hypothesis testing task, if/only thinking, and outcome bias. Composite, All Tasks = rational thinking composite score of performance on all seven tasks in the replication and extension experiment.

roughly equal in magnitude (.347 to .394). All were statistically significant. The remaining correlations in the table are the results from a replication and extension experiment (Study 2 of Stanovich & West, 1998c). These correlations indicate that the correlations involving the syllogistic reasoning task, statistical reasoning task, and argument evaluation task were similar in magnitude to those obtained in Study 1. The correlations involving the four new tasks were also all statistically significant. The sign on the hypothesis testing, outcome bias, and if/only thinking tasks was negative because high scores on these tasks reflect susceptibility to non-normative cognitive biases. The correlations on the four new tasks were generally lower (range .172 to .239) than the correlations involving the other tasks (.371 to .410). However, it must again be emphasized that the logistical constraints dictated that the scores on some of the new tasks were based on an extremely small sample of behavior. The outcome bias score was based on only a single comparison and the if/only thinking score was based on only two items.

The remaining correlations in Table 14.2 concern composite variables. The first composite involved the three tasks that were carried over from the previous experiment – the syllogistic reasoning, statistical reasoning, and argument evaluation tasks. The scores on each of these three tasks were standardized and summed to yield a composite score. The composite's correlation with SAT scores was .530. A second composite was formed by summing the standard scores of the remaining four tasks: covariation judgment, hypothesis testing, outcome bias, and if/only thinking (the latter three scores reflected so that higher scores represent more normatively correct reasoning). SAT Total scores displayed a correlation of .383 with this composite. Finally, both of the composites were combined into a composite variable reflecting performance on all seven tasks and this composite displayed a correlation of .547 with SAT scores.

So there is no question that cognitive ability is implicated in performance on many of these reasoning tasks and to some extent may be accounting for the domain generality of performance across the tasks. One further question that we have addressed is whether the residual variance (after cognitive ability is accounted for) is systematic or whether instead it appears to be error variance. Such a model – to be elaborated shortly – could preserve a Parglossian assumption of perfect human rationality (see Stanovich, 1999; Stanovich & West, 2000b). Differences in algorithmic-level computational capacity would explain part of the discrepancy between normative and descriptive models of behavior, and any remaining deviations from the prescriptive could be attributed to performance errors. There would be no need to posit a systematically suboptimal cognitive functioning at the intentional level of analysis. In contrast, evidence that the residual variance (after partialing cognitive ability) was systematic would mean that not all of the normative/descriptive gap could be attributed to computational limitations and performance errors, and it would support the idea that the intentional-level model of behavior is characterized by systematic suboptimal functioning.

The data in Table 14.3 are relevant to this issue. The correlations from the four tasks in Study 1 of Stanovich and West (1998c) are recapitulated below the diagonal. What is different in this table is that above the diagonal are

**TABLE 14.3. Inter correlations Among Several Reasoning Tasks**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Syllogisms</td>
<td>.257*</td>
<td>.222*</td>
<td>.213*</td>
<td></td>
</tr>
<tr>
<td>2. Selection task</td>
<td>.363**</td>
<td>.150</td>
<td>.215**</td>
<td></td>
</tr>
<tr>
<td>3. Statistical reasoning</td>
<td>.314**</td>
<td>.258**</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td>4. Argument Evaluation</td>
<td>.349**</td>
<td>.310**</td>
<td>.117</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Data from Study 1 of Stanovich & West, 1998c.*

*Note: Zero-order correlations are below the diagonal and correlations with SAT Total score partialled out are presented above the diagonal.

* = p < .05; ** = p < .01; *** = p < .001, all two-tailed; N = 188 to 195.
presented the correlations among the tasks after differences in cognitive ability have been statistically removed. The results indicate that four of the six associations were still significant even after variance in SAT scores had been partialed out.

A similar analysis was carried out on the composite rational thinking variables from Study 2 of Stanovich and West (1998c) displayed at the bottom of Table 14.2 and described previously. The two composite scores—despite being composed of vastly different reasoning tasks—displayed a significant correlation of .395 (p < .001, n = 546) this is another way of capturing the domain generality in normative responding that we discussed earlier. However, both Composite #1 and Composite #2 displayed significant correlations with SAT scores (.530 and .383, respectively, both ps < .001). Thus again the possibility remains that the association between the two thinking indices derives from their common association with cognitive ability. However, the correlation between Composite #1 and Composite #2 when the variance due to SAT scores was partialed out remained significant (partial r = .242, F(1, 526) = 32.82, p < .001). In summary, the two separate multivariate studies presented here (as well as others that we have conducted) both produced evidence indicating that there is systematic variance in various reasoning tasks that is not explained by variation in cognitive ability.3

There are actually two different ways to examine whether computational limitations plus random performance errors explain all of the variance in performance on these tasks. The first is that which we have just applied—to determine the covariance among reasoning tasks after cognitive ability has been partialed out. The second is to examine whether there are personality variables—that is, thinking dispositions of the type that have been studied in the critical thinking literature—that can explain the variation in performance after cognitive ability has been accounted for. We have examined this question as well.

The latter analyses are framed by an assumption that often is unarticulated in the psychological literature: that cognitive abilities and thinking dispositions are constructs at different levels of analysis in a cognitive theory (Anderson, 1990; Dennett, 1987; Marr, 1982; Newell, 1982; Stanovich, 1999) and that they may do separate explanatory work in a descriptive theory of human reasoning performance. Specifically, each level of analysis in cognitive theory frames a somewhat different issue. At the biological level the paramount issue is whether the physical mechanism has the potential to instantiate certain complex algorithms. At the algorithmic level, the key issue is one of computational efficiency. In contrast, it is at the intentional level that issues of rationality arise. Omnibus measures of cognitive capacities, such as intelligence tests, index individual differences in the efficiency of processing at the algorithmic level. In contrast, thinking dispositions as traditionally studied in psychology (Cacioppo, Petty, Feinstein, &

<table>
<thead>
<tr>
<th>TABLE 14.4: Correlations between the Reasoning Tasks and SAT Total Score and Thinking Dispositions Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Argument evaluation task</td>
</tr>
<tr>
<td>Syllogisms</td>
</tr>
<tr>
<td>Statistical reasoning</td>
</tr>
<tr>
<td>Covariation: Δp</td>
</tr>
<tr>
<td>Hypothesis testing bias</td>
</tr>
<tr>
<td>Outcome bias</td>
</tr>
<tr>
<td>If/Only thinking</td>
</tr>
<tr>
<td>RT1 Composite</td>
</tr>
<tr>
<td>RT2 Composite</td>
</tr>
<tr>
<td>RT Composite, All Tasks</td>
</tr>
</tbody>
</table>

Source: Data from Study 2 of Stanovich & West, 1998c.
Note: TDC = Thinking dispositions composite score. Composite #1 = standard score composite of performance on argument evaluation task, syllogisms, and statistical reasoning. Composite #2 = standard score composite of performance on covariation judgment, hypothesis testing task, if/only thinking, and outcome bias. Composite, All Tasks = rational thinking composite score of performance on all seven tasks.

Jarvis, 1996; Kardas & Scholes, 1996; Klaczynski et al., 1997; Kruglanski & Webster, 1996; Schommer, 1990, 1993, 1994; Stanovich & West, 1997; Sternberg, 1997) index individual differences at the intentional level of analysis. They are telling us about the individual’s goals and epistemic values—and they are indexing broad tendencies of pragmatic and epistemic self-regulation. If thinking dispositions correlate with individual differences residualized on cognitive ability, then this will be an additional indication that variance in reasoning performance is caused by actual differences in intentional psychology.

The left column of figures in Table 14.4 recapitulates the results regarding cognitive ability from Study 2 of Stanovich and West (1998c) that were discussed previously. The second column of figures in this table includes the correlations with a composite measure of thinking dispositions that was employed in that study. This measure is the additive combination of several questionnaire subscales designed to tap epistemic self-regulation (Goldman, 1986; Harman, 1995; Nozick, 1993; Thagard, 1992). The subscales overrepresented dispositions with potential epistemic significance, for example: “the disposition to weigh new evidence against a favored belief heavily (or lightly), the disposition to spend a great deal of time (or very little) on a problem before giving up, or the disposition to weigh heavily the opinions of others in forming one’s own” (Baron, 1985, p. 15).
Baron (1985, 1988) has called such tendencies dispositional rather than actively open-minded thinking. Overall, the subscales in this thinking dispositions questionnaire measure the following dimensions: epistemological absolutism, willingness to perspective-switch, willingness to decontextualize, and the tendency to consider alternative opinions and evidence (see also Sá et al., 1999; Schommer & Walker, 1995; Stanovich & West, 1997; Webster & Kruglanski, 1994).

The second column of correlations in Table 14.4 indicates that the thinking dispositions composite score (TDC) displayed significant correlations with each of the tasks in the Stanovich and West (1998c) study. In each case, the direction of the relationship was the same as that observed for cognitive ability. In most cases, the correlations involving the TDC were lower than those involving the SAT, but in several cases (outcome bias, if/only thinking) the magnitude of the correlations was similar. The correlations between the TDC and the three composite indices of reasoning were moderate in size (.413, .324, and .442, respectively).

There appears to be no question that there are consistent and replicable relationships between intention-level-thinking dispositions and performance on a variety of tasks from the reasoning and heuristics and biases literature. To find out whether the thinking dispositions explained unique variance after cognitive ability was controlled, a composite score was computed combining performance on all seven tasks. SAT total scores and the TDC attained a multiple R with this criterion variable of .627 (F(2, 526) = 170.56, p < .001). SAT total was a significant unique predictor (partial correlation = .496, unique variance explained = .198, p < .001) as was the TDC (partial correlation = .366, unique variance explained = .094, p < .001). Thus, there were consistent indications in the data of Stanovich and West (1998c) and in other studies (Sá et al., 1999; Stanovich & West, 1997) that thinking dispositions do in fact explain variance on a variety of reasoning and decision-making tasks after the variance in cognitive ability has been accounted for. The residual variance does appear to be systematic and predictable. These results refute the notion that all performance variability can be accounted for by algorithmic-level limitations and performance errors. Indicators of intention-level epistemic attitudes are consistent unique predictors of normative response tendencies. And with this conclusion we have engaged our third issue.

**Issue 3: Individual Differences in Reasoning: Implications for the Rationality Debate?**

The interpretation of the gap between descriptive models and normative models in the human reasoning and decision-making literature has been the subject of contentious debate for almost two decades now (see Baron, 1994; Cohen, 1981, 1983; Evans & Over, 1996; Gigerenzer, 1996; Kahneman, 1981; Kahneman & Tversky, 1983, 1996; Koehler, 1996; Samuels, Stich, & Tremoulet, 1999; Stein, 1996; Vranas, 2000). The debate has arisen because some investigators wished to interpret the gap as indicating that human cognition was characterized by systematic irrationalities. In our “Behavioral and Brain Sciences” target article on these debates (Stanovich & West, 2000b), these investigators were labeled the Moralists, due to the emphasis that they place on reasoning about human cognition. Disputing the attribution of irrationality were numerous investigators (who were termed Panglossians in this discussion, see also Stanovich, 1999) who argued that there were other reasons why reasoning might not accord with normative theory (see Cohen, 1981 and Stein, 1996 for extensive discussions of the various possibilities – reasons that prevent the ascension of irrationality to subjects).

Four alternative interpretations for the normative/descriptive gap were extensively discussed in Cohen’s (1981) classic “Behavioral and Brain Sciences” article and in Stein’s (1996) more recent book-length treatment of the rationality debate. First, reasoning might depart from normative standards due to performance errors – temporary lapses of attention, memory deactivation, and other sporadic information processing mishaps. Second, there may be stable and inherent computational limitations that prevent the normative response (Cherniak, 1986; Goldman, 1978; Harman, 1995; Oaksford & Chater, 1993, 1995, 1998; Stich, 1990). Third, in interpreting performance, we might be applying the wrong normative model to the task (Koehler, 1996; Vranas, 2000). Alternatively, we may be applying the correct normative model to the problem as set, but the subject might have construed the problem differently and be providing the normatively appropriate answer to a different problem (Adler, 1984, 1991; Berkeley & Humphreys, 1982; Broome, 1990; Hilton, 1995; Schwarz, 1996).

The data discussed in the earlier sections of this chapter relate to the first two of these alternative explanations. We have seen that several classic tasks from the reasoning and decision-making literature seem to have moderate computational limitations that might prevent some subjects from deriving the correct response – and, of course, performance errors are a factor in responding. However, Stanovich (1999; Stanovich & West, 2000b) defined a position – termed the Apologist position – which posited that the first two factors, performance errors and computational limitations, could explain all of the normative/descriptive gap. Such a view predicts that once capacity limitations have been controlled, the remaining variations from normative responding will be essentially unpredictable (all being due to performance errors). In contrast, we have just shown in the previous section that there is significant covariance among the residualized scores from a variety of tasks. The residual variance was also systematically associated with thinking dispositions that were conceptualized as characteristic intention-level attitudes reflecting epistemic regulation. That cognitive/personality variables can explain normative/descriptive...
discrepancies that remain after computational limitations have been accounted for signals a systematically suboptimal intentional-level model of performance.

Thus, the Apologist’s position does have a substantial grain of truth, but in the extreme cannot account for all the reliable individual differences. However, the Panglossian position does have two additional arguments to use in an attempt to completely close the normative/descriptive gap. First, the possibility of incorrect norm application arises because any particular laboratory problem must be matched to an appropriate normative model. Matching a problem to a normative model is rarely an automatic or clear cut procedure. The complexities involved in matching problems to norms make possible the argument that the gap between the descriptive and normative occurs because psychologists are applying the wrong normative model to the situation. It is a potent strategy for the Panglossian theorist to use against the advocate of Meliorism and such claims have become quite common in critiques of the heuristics and biases literature (see Cohen, 1981; Gigerenzer, 1996; Stanovich, 1999; Stein, 1996; Vranas, 2000).

A second possibility is the argument that although the experimenter may well be applying the correct normative model to the problem as set, the subject might be construing the problem differently and be providing the normatively appropriate answer to a different problem — in short, that subjects have a different interpretation of the task (see, for example, Adler, 1984; 1993; Broome, 1990; Henle, 1962; Hilton, 1993; Levinson, 1995; Margolis, 1987; Schick, 1987; 1997; Schwartz, 1996). As with incorrect norm application, the alternative construal argument locates the problem with the experimenter. However, it is different in that in the wrong norm explanation it is assumed that the subject is interpreting the task as the experimenter intended — but the experimenter is not using the right criteria to evaluate performance. In contrast, the alternative task construal argument allows that the experimenter may be applying the correct normative model to the problem the experimenter intends the subject to solve — but posits that the subject has construed the problem in some other way and is providing a normatively appropriate answer to a different problem.

How to decide whether experimenters are applying the appropriate normative model to a task and assuming the right task construal is a vexing problem. We have followed the lead of the Panglossian theorists here in consciously committing the naturalistic fallacy to get a foothold on this issue. What we are referring to is, to paraphrase Stein’s (1996) terminology, the way the so-called “reject-the-experimenter” strategy is used by Panglossian theorists when the modal response on a task departs from the response deemed normative by the experimenter. It is noteworthy that this strategy is exclusively used to eliminate gaps between descriptive models of performance and normative models — although this connection is not a necessary one. When this type of critique is employed, the normative model or task construal that is suggested as a substitute for the one traditionally used is one that coincides perfectly with the descriptive model of the subjects’ performance — thus preserving a view of human rationality as ideal. It is rarely noted that the strategy could be used in just the opposite way — to create gaps between the normative and descriptive. Situations where the modal response coincides with the standard normative model could be critiqued, and alternative models could be suggested that would result in a new normative/descriptive gap. But this is never done. The Panglossian camp, often highly critical of empirical psychologists, is never critical of psychologists who design reasoning tasks in instances where the modal subject gives the response the experimenters deem correct. Ironically, in these cases, according to the Panglossians, the same psychologists seem never to err in their task designs and interpretations.

There, perhaps, is some logic to this bias displayed by Panglossians. Might not the majority of subjects be telling us something — something that the experimentalists missed? Hence our characterization of this attitude as a case of consciously committing the naturalistic fallacy — in our search for what is good reasoning, letting the subjects themselves tell us. What this means is that in an important sense the norms being endorsed by the Panglossian camp are conditioned by descriptive facts about human behavior. The rationality debate itself is, reflexively, evidence that the descriptive models of actual behavior condition expert notions of the normative. That is, there would have been no debate (or at least much less of one) had people behaved in accord with the then-accepted norms.

But if we are going to use descriptive facts about behavior to give us clues as to what are the proper norms to apply and construals to consider reasonable, we must ask why the modal response is the only aspect of group performance that is relevant? Might the pattern of responses around the mode tell us anything? Or do the moments of the distribution contain no normative information? And finally, what about the the rich covariance patterns that would be present in any multivariate experiment? Are these totally superfluous — all norm-relevant behavioral information residing in the mode? We think not.

PUTTING INDIVIDUAL DIFFERENCES TO WORK IN THE RATIONALITY DEBATE: THE UNDERSTANDING/ACCEPTANCE PRINCIPLE

One goal of the present research program is to expand the scope of the descriptive information used to condition our views about appropriate norms and task construals by using Spearman’s positive manifold as a diagnostic tool. Larrick, Nisbett, and Morgan (1993) made just such an argument in their analysis of what justified the cost-benefit reasoning of microeconomics. Their point was that: “Intelligent people would be more likely to
use cost-benefit reasoning. Because intelligence is generally regarded as being the set of psychological properties that makes for effectiveness across environments... intelligent people should be more likely to use the most effective reasoning strategies than should less intelligent people" (p. 333). Larrick et al. (1993) are alluding to the fact that we may want to condition our inferences about appropriate norms based not only on what response the majority of people make but also on what response the most cognitively competent subjects make.

We traced this basic idea, which we termed the understanding/acceptance principle (see Stanovich, 1999; Stanovich & West, 1999, 2000b) to a 25-year-old paper by Slovic and Tversky (1974) in which in an imaginary dialog between Allais and Savage it was argued that it should be the case that "the deeper the understanding of the axiom, the greater the readiness to accept it" (pp. 372-373). Thus, a positive correlation between understanding and acceptance would suggest that the gap between the descriptive and normative was due to an initial failure to fully process and/or understand the task. The basic point is that a normative/descriptive gap that is disproportionately created by subjects with a superficial understanding of the problem provides no warrant for amending the models applied and task construals assumed by the designers of the problem.

There are two generic strategies for applying the understanding/acceptance principle based on the fact that variation in understanding can be created or it can be studied by examining naturally occurring individual differences. Slovic and Tversky (1974) employed the former strategy by providing subjects with explicated arguments supporting the Allais or Savage normative interpretation (see also Doherty, Schiavo, Tveney, & Mynatt, 1981; Stanovich & West, 1999). As an alternative to manipulating understanding, the understanding/acceptance principle can be transformed into an individual differences prediction. For example, the principle might be interpreted as indicating that more reflective, engaged, and intelligent reasoners are more likely to respond in accord with normative principles. Thus, it might be expected that those individuals with cognitive/personality characteristics more conducive to deeper understanding would be more accepting of the appropriate normative principles for a particular problem. This was the emphasis of Larrick et al. (1993) when they argued that more intelligent people should be more likely to use cost-benefit principles. Similarly, need for cognition – a dispositional variable reflecting the tendency toward thoughtful analysis and reflective thinking – has been associated with aspects of epistemic and practical rationality (Cacioppo, et al., 1996; Kardasz & Soles, 1996; Klaczynski et al., 1997; Smith & Levin, 1996; Stanovich & West, 1999; Verplanken, 1993).

In fact, it is probably helpful to articulate the understanding/acceptance principle somewhat more formally in terms of Spearman’s positive manifold – the fact that different measures of cognitive ability almost always correlate with each other (see Carroll, 1993, 1997). The individual differences version of the understanding/acceptance principle puts positive manifold to use in areas of cognitive psychology where the nature of the appropriate normative model to apply is in dispute. The point is that scoring a vocabulary item on a cognitive ability test and scoring a probabilistic reasoning response on a task from the heuristics and biases literature are not the same. The correct response in the former task has a canonical interpretation agreed upon by all investigators; whereas the normative appropriateness of responses on tasks from the latter domain has been the subject of extremely contentious dispute (Cohen, 1981, 1982, 1986; Cosmides & Tooby, 1996; Einhorn & Hogarth, 1981; Gigerenzer, 1991, 1993, 1996; Kahneman & Tversky, 1996; Koehler, 1996; Stein, 1996). Positive manifold between the two classes of task would only be expected if the normative model being used for directional scoring of the tasks in the latter domain is correct.

Oaksford and Sellen (2000) have emphasized this point by arguing that "in studying reasoning it is important to bear in mind that unlike tasks in almost any other area of cognition, reasoning tasks do not come pre-stamped with the 'correct' answer. The correct answer has to be discovered because it depends on how people interpret the task (Oaksford & Chater, 1993, 1995, 1998). In this respect, paradoxical individual differences, where a dysfunctional trait correlates with some preconceived notion of the correct answer, are particularly compelling" (p. 692). Thus, an important part of our method was to use the very obviousness of positive manifold as a diagnostic tool. If positive manifold is indeed to be expected, then another observation of it (while unsurprising in and of itself) might then be viewed as converging evidence for the normative model applied and/or task construal assumed in the scoring of the problem. Conversely, violations of positive manifold might be thought to call into question the normative model and task construal used to score the problem.

It is important to emphasize that we are proposing here to use patterns of covariation to help to determine whether it is appropriate to apply a particular normative model to a specific problem rather than to justify particular normative models themselves. This is a distinction that we were forced to clarify in our reply to the Behavioral and Brain Sciences commentators (Stanovich & West, 2000b). We agreed with several commentators that such individual differences patterns do not bear on the validity of normative models themselves but instead relate to disputes about which normative model should be applied to a particular experimental situation. There are many cases of the latter situation that have heretofore resisted resolution by argument (the heuristics and biases literature is littered with examples; refer to the commentary on Cohen, 1981, and Kahneman & Tversky, 1996).

With this caveat in mind, it is thus interesting to note that the direction of all of the correlations displayed in Tables 14.1–4 is consistent with
the standard normative models used by psychologists. The tasks – as traditionally scored – correlate among themselves and also with general intelligence. The directionality of the systematic correlations with intelligence is embarrassing for critics who question the appropriateness of the norms applied and construals assumed by the designers of these tasks. Surely we generally would want to avoid the conclusion that individuals with more computational power are systematically computing the nonnormative response. Such an outcome would be nearly unique in a psychometric field that is one hundred years and thousands of studies old (Brody, 1997; Carroll, 1993, 1997; Lubinski & Humphreys, 1997; Neisser et al., 1996; Sternberg & Kaufman, 1998). It would mean that Spearman’s (1904, 1927) positive manifold for cognitive tasks – virtually unchallenged for 100 years (but see Sternberg et al., 2001) – had finally broken down. Obviously, parsimony dictates that positive manifold remains a fact of life for cognitive tasks and that the normative model originally assumed appropriate actually is.

However, an application of the understanding/acceptance principle does not always work out this way. We have occasionally observed negative correlations between cognitive ability and the traditional scoring of a task. Again, given that positive manifold is the norm for cognitive tasks, such negative correlations might be taken as signals that the wrong normative model is being applied or that there are alternative construals of the task that are more appropriate.

For example, in several experiments (Stanovich & West, 1998c, 1999), we have examined some of the noncausal base-rate problems (those involving base rates with no obvious causal relationship to the criterion behavior) that are notorious for provoking philosophical dispute. One was the infamous cab problem studied by Bar-Hillel (1980) and Tversky and Kahneman (1982). Using several versions of this problem (Stanovich & West, 1998c, 1999), we have never found a tendency for the Bayesian responders to have higher cognitive ability. In some experiments there have been no significant differences found, and in others we have even found that the Bayesian group actually had significantly lower SAT scores and/or scored significantly lower in need for cognition. So an application of the understanding/acceptance principle can support the Panglossian position by showing that more intelligent and reflective subjects sometimes differentially reject the task construals and normative models applied to tasks by the designers of the problems in the reasoning and decision-making literature.

Most important for present purposes, however, is not which tasks go in which direction, nor how many tasks do, but to simply illustrate how a particular use of individual differences may have implications for debates about theories of the gap between normative models and descriptive models of reasoning performance. In reply to Cohen’s (1981) well-known critique of the assumptions in the reasoning literature – surely the most often cited of such critiques – Jepson, Krantz, and Nisbett (1983) argued that “Cohen postulates far too broad a communality in the reasoning processes of the ‘untutored’ adult” (p. 495). Jepson et al., we argue, were right on the mark, but their argument has been largely ignored in more recent debates. For example, philosopher Nicholas Rescher (1988) argues that “to construe the data of these interesting experimental studies . . . to mean that people are systematically programmed to fallacious processes of reasoning – rather than merely that they are inclined to a variety of (occasionally questionable) substantive suppositions – is a very questionable step” (p. 196). There are two parts to Rescher’s (1988) point here: the “systematically programmed” part and the “inclination toward questionable suppositions” part. Rescher’s (1988) focus, like that of many who have dealt with the philosophical implications of the idea of human irrationality (Cohen, 1981, 1982, 1986; Davidson, 1980; Dennett, 1987; Goldman, 1986; Harman, 1995; Kornblith, 1993; Stein, 1996; Stich, 1990), is on the issue of how humans are systematically programmed. “Inclinations toward questionable suppositions” are only of interest to those in the philosophical debates as mechanisms that allow one to drive a wedge between competence and performance (Cohen, 1981, 1982; Rescher, 1988) – thus maintaining a theory of near-optimal human rational competence in the face of a host of responses that seemingly defy explanation in terms of standard normative models.

Analogously to Rescher, Cohen (1982) argues that there really are only two factors affecting performance on rational thinking tasks: “normatively correct mechanisms on the one side, and adventitious causes of error on the other” (p. 252). Not surprisingly given such a conceptualization, the processes contributing to error (“adventitious causes”) are of little interest to Cohen (1981, 1982). Human performance arises from an intrinsic human competence that is impeccably rational, but responses occasionally deviate from normative correctness due to inattention, memory lapses, lack of motivation, and other fluctuating but basically unimportant causes. There is nothing in such a view that would motivate any interest in patterns of errors or individual differences in such errors.

One of the purposes of the present research program is to reverse the figure and ground in the rationality debate, which has tended to be dominated by the particular way that philosophers frame the competence/performance distinction. From a psychological standpoint, there may be important implications in precisely the aspects of performance that have been backgrounded in this controversy ("adventitious causes," "peccadillos"). That is, whatever the outcome of the disputes about how humans are “systematically programmed” (Cosmides & Tooby, 1996; Johnson-Laird & Byrne, 1994, 1993; Johnson-Laird, Byrne, & Schaeken, 1994; Oakford & Chater, 1994, 1996; O'Brien, Braine, & Yang, 1994; Rips, 1994), variation in the "inclination toward questionable suppositions" is of psychological
interest as a topic of study in its own right. The experiments discussed in this chapter provide at least tentative indications that the “inclination toward questionable suppositions” has some degree of domain generality and that it is predicted by thinking dispositions that concern the epistemic and pragmatic goals of the individual and that are part of people’s intentional-level psychology.

NEW DIRECTIONS FOR RESEARCH ON INDIVIDUAL DIFFERENCES IN REASONING: THE PRAGMATIC CALIBRATION OF KNOWLEDGE ACQUISITION

In the research program just described, we have applied our individual differences analyses to normative issues surrounding a variety of different types of rationality. That is, philosophers and decision scientists have distinguished many different concepts of rationality—and most have been addressed by at least one task included in our research program. Perhaps the clearest and most well-known distinction is that between rationality of belief and rationality of action (Audi, 1993a, 1993b; Harman, 1995; Nozick, 1993; Pollock, 1995; Sloman, 1999). The rationality of belief—how accurately a person’s belief network represents the external world—has been variously termed theoretical rationality, evidential rationality, or epistemic rationality (Audi, 1993b; Foley, 1987; Harman, 1995). The rationality of action—how well a person’s actions maximize the satisfaction of their desires, given their beliefs—has been variously termed practical, pragmatic, instrumental, or means/ends rationality (Audi, 1993b; Harman, 1995; Nathanson, 1994; Nozick, 1993).

Although the distinction between theoretical rationality and practical rationality is the most common distinction made in the philosophical literature, cognitive psychologists have discussed related distinctions and terms. For example, Kahneman (2000) has distinguished “the ability to reason correctly about immediately available information” (p. 682) from the “decision theory rationality [that] is defined by the coherence of beliefs and preferences” (p. 682). Alternatively, Evans and Over (1996) make the distinction between rationality, and rationality_2, the former defined as reasoning and acting “in a way that is generally reliable and efficient for achieving one’s goals” (p. 8). Thus, rationality, sounds much like instrumental rationality, as traditionally defined. Rationality_2, in contrast, refers to reasoning and acting “when one has a reason for what one does sanctioned by a normative theory” (p. 8). The distinction drawn by Evans and Over (1996) brings to the fore the mechanism used to pursue personal goals (mechanisms of conscious, reason-based rule-following versus tacit heuristics).

We are currently examining individual differences in an aspect of rationality that is not captured by any of these previous distinctions. This is because this particular aspect of rationality is a peculiar amalgamation of epistemic and practical rationality as traditionally defined. In a nuanced discussion of the relation between epistemic and practical rationality, Foley (1991) discusses the arguments of the so-called “evidentialists” who argue that there cannot be good non-evidential reasons for belief. One argument of the evidentialists is that non-evidential reasons would be redundant because there is one overriding practical concern that directs epistemic cognitive operations, to “maintain a comprehensive stock of beliefs that contains few false beliefs” (p. 372). The reason for this is that we are constantly faced with a plethora of decision situations that cannot be anticipated in advance. Many of these require an online response under time constraints that do not allow for an additional period of information search and knowledge acquisition. Thus, without the time to collect more information, and not knowing what the decision will be in advance, an individual has to deal with the situation based on the stock of information already collected. The stock of beliefs that is most likely to foster good on-the-spot decision making is a stock of beliefs that is large and accurate. Thus, the argument goes that “you can usually ignore practical reasons in your deliberations about what to believe. You can do so because ordinarily these practical reasons simply instruct you to acquire beliefs for which you have good evidence” (p. 373).

But Pollock (1995) has argued that the interplay of epistemic and practical rationality plays out in ways that make the degree of rational justification for belief subject to normative evaluation based on practical concerns. He concocts the story of a ship’s captain on a busman’s holiday on a Caribbean cruise liner. The ship seems well equipped and the captain casually notes the lifeboats, wonders how many there are, and consults a cheap brochure given to all passengers. However, later on, an accident occurs, disabling the entire ship’s crew and putting the ship in danger of sinking. The vacationing captain is now in charge of the cruise liner and immediately wants to know whether there are enough lifeboats on board. His belief based on the holiday brochure is no longer sufficient—he must have the lifeboats counted carefully and accurately. The changed decision pragmatics have implications for how we evaluate the proper justification for belief. The evaluation of whether the degree of justification was well calibrated depends on the importance of the situation. More generally, as Pollock (1995) notes, “the degree of interest in a question determines the degree of justification a rational agent must have for an answer” (p. 49).

We propose that this argument of Pollock’s (1995) can be scaled up into a more general argument about knowledge acquisition. Consider a real-life case that is a problematic one to analyze using standard categories of types of rationality. Several years ago, in the mid-1990s, one of this chapter’s authors heard a television report that the National Highway Safety Administration had determined that something like 40% (the actual
percentage is immaterial to our example) of small children in America were still riding in automobiles without being secured with seatbelt restraints. How are we to interpret this appalling statistic from the standpoint of evaluating the rationality of the parents of those children? Starting with a Panglossian bias to preserve the rationality of the parents, we could call a certain percentage of these cases "performance errors" — some of these are children who are regularly belted but whose parents forgot on a certain occasion.

But highway safety commissions can assure us that this is not the entire story — some children are unbelted time after time. Their parents' behavior is not a performance error because it is systematic. How do we preserve the rationality of the parents in these cases? We are not prone to take one escape hatch that is available to the Panglossian — that perhaps the parents really do not value their children. What creates the paradox is that the parents' behavior (they are putting their children at great risk by not securing them with seatbelts) is at total odds with their desires and goals (they love their children and desire to protect them). One way out of the paradox is to deny the latter — like the Panglossian economist who replies to a person who claims to like widgets but does not purchase them that, in fact, they do not really like widgets. This assumption preserves the Panglossian default of perfect rationality but at the expense of a dim view of the characteristics of our fellow humans. Most prefer not to escape the paradox by going in this direction.

Instead, the more popular way to resolve the seeming paradox is to retreat to the extreme strictures of what Elster (1985) calls a thin theory of rationality — where beliefs are treated as fixed and not subject to evaluation and, likewise, the content of desires is not evaluated. That is, the individual's goals and beliefs are accepted as they are, and debate centers only on whether individuals are optimally satisfying desires given beliefs. The thin theory — plus a Panglossian default — then simply says that, in this case, given that the desire is fixed (these parents love their children) and the behavior is fixed (they did not, in fact, secure the children with seatbelts), then it must follow that the belief behind the action must have been incorrect. The parents must not have known that unbelted children in automobile collisions — the leading cause of childhood death (National Highway Traffic Safety Administration, 1999) — are in unusual danger.

But is this really a satisfactory solution? Are we really assuaged that there is nothing wrong here? We submit that there are still grounds for worry. If we merely slip the Panglossian blinders, we will see that this example points to an entire domain of rationality that is open to assessment — the calibration of knowledge acquisition according to practical goals. We submit that the seatbelt example does reveal a failure of rational thinking that is related to notions of epistemic responsibility that philosophers have discussed (e.g., Code, 1987). Specifically, it is fine to say, as the thin theory does, that the parents were not irrational because they did not know that unbelted children are in particular danger. But the issue we are raising here is that perhaps it is also appropriate to ask the question: Why didn't they know? The media has been saturated with seatbelt warnings for over two decades now. Educational efforts in schools and communities are directed toward it. Of course, it is a key component of all driver training courses. Information about the importance of seatbelts for children is not hard to acquire.

Just as in the Pollock (1995) ship captain example, there seems to be a requirement of calibration here. It was important for the ship's captain to calibrate his knowledge of the lifeboats to his situation (was he a mere passenger or was he in charge of the boat). Similarly, over a person's lifetime, it is critical to acquire knowledge in domains that are most relevant to fulfilling one's most important goals. This is what we are calling the pragmatic calibration of knowledge acquisition. The knowledge that an individual acquires is unevenly distributed, as is the importance of all possible activities in the person's life. As Foley (1991) notes, true beliefs in a given domain are most likely to foster goal achievement in that domain. But our goals are not evenly spread over domains. The distribution of true beliefs should thus be in line with the distribution of importance across the various domains of life.

Knowledge acquisition is effortful and the cognitive resources for it are limited. There is a limited amount of time and effort to spend in epistemic activities, and it is important that that effort be calibrated so that it is directed at knowledge domains that are connected to goals we deem important. If we say that something is of paramount importance to us (our children's safety, for example), then it is incumbent on us to know something about these things we deem of such importance. The parents of the unbelted children are irrational on this view — not from the peculiar slice-in-time perspective of the thin theory (with its emphasis on a particular instance where they did not belt their children because they "didn't know" they should) — but from a long-term perspective. They have instead poorly calibrated their knowledge acquisition over a long time period of their lives — they have a systemic problem in the domain of rational thought and cognitive activity. Individual differences in this new type of rationality are currently under investigation in our laboratory.

Notes

1. The Panglossian position is represented by philosophers and cognitive scientists who see no gaps between the descriptive and normative models of performance. The Panglossian feels that people reason as well as anyone could possibly reason. Observed discrepancies are attributed to transitory performance errors, the wrong norm being applied by the experimenter, or an alternative
construal of the task on the part of the subject (see Stanovich, 1999, pp. 4-9, for a discussion).

2. Following Dennett (1987) and the taxonomy of Anderson (1990), we distinguish the algorithmic/design level from the rational/intentional level of analysis in cognitive science. The latter provides a specification of the goals of the system’s computations (what the system is attempting to compute and why). At this level, we are concerned with the goals, system beliefs relevant to those goals, and the choice of action that is rational given the system’s goals and beliefs (see Stanovich, 1999, pp. 9-12, for a discussion).

3. Although we operationalized cognitive ability here using SAT total scores, separating out the verbal and mathematical scores on the SAT reveals essentially identical trends. Additionally, in many of our studies we have employed converging measures of cognitive ability - an additional test of crystallized intelligence such as a vocabulary measure; a measure of fluid intelligence such as Raven matrices, or a working memory task. The correlations obtained with these measures generally converge with those obtained with the SAT. Finally, there are good reasons to view such cognitive ability tests as crude indicators of the overall level of current computational efficiency (see Stanovich & West, 2000a, pp. 704-705, for a discussion).

References


Individual Differences in Reasoning


Individual Differences in Reasoning


Individual Differences in Reasoning


