# 17

# *Individual Differences in Reasoning and the Heuristics and Biases Debate*

*KEITH E. STANOVICH and RICHARD F. WEST*

A substantial research literature — one comprising literally hundreds of empirical studies conducted over nearly three decades — has firmly established that people's responses often deviate from the performance considered normative on many reasoning tasks. For example, people assess probabilities incorrectly, display confirmation bias, test hypotheses inefficiently, violate the axioms of utility theory, do not properly calibrate degrees of belief, and display numerous other information-processing biases (for summaries of the large literature, see Baron, 1994; Evans & Over, 1996; Kahneman & Tversky, 1996; Osherson, 1995; Shafir & Tversky, 1995). Indeed, demonstrating that descriptive accounts of human behavior diverged from normative models was a main theme of the so-called heuristics and biases literature of the 1970s and early 1980s (see Kahneman, Slovic, & Tversky, 1982).

The interpretation of the gap between the descriptive and the normative in the human reasoning and decision-making literature has been the subject of contentious debate for almost two decades now (for summaries, see Cohen, 1981; Evans & Over, 1996; Gigerenzer, 1996; Kahneman & Tversky, 1996; Stanovich, in press; Stein, 1996). The debate has arisen because some investigators wished to interpret the gap between the descriptive and the normative as indicating that human cognition was characterized by systematic irrationalities. Disputing this contention were numerous investigators who argued that there were other reasons why instances of actual reasoning performance might not accord with the normative theory (see Cohen, 1981, and Stein, 1996, for extensive discussions of the various possibilities). First, perfor-

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Keith E. Stanovich and a James Madison University Program Faculty Assistance Grant to Richard F. West. We thank Penny Chiappe, Alexandra Gottardo, Walter Sa, Robin Sidhu, and Ron Stringer for their assistance in data coding.

mance may depart from normative standards because of performance errors: temporary lapses of attention, memory deactivation, and other sporadic information-processing mishaps. Second, there may be computational limitations that prevent the normative response (Cherniak, 1986; Oaksford & Chater, 1993, 1995). Third, in interpreting performance, researchers might be applying the wrong normative model to the task. Alternatively, the correct normative model may be applied to the problem as set, but the participant might have construed the problem differently and may have provided the normatively appropriate answer to a different problem (Adler, 1984; Hilton, 1995; Schwarz, 1996).

One aspect of performance that has been neglected by all parties in these disputes has been individual differences. What has largely been ignored is that although the average person in these experiments might well display an overconfidence effect, underutilize base rates, choose *P* and *Q* in the selection task, commit the conjunction fallacy, and so forth, on each of these tasks, some people give the standard normative response. In a series of studies, our research group has been exploring the possibility that these individual differences and their patterns of covariance might have implications for explanations of why human behavior often departs from normative models. In this chapter, we illustrate how we have explored these implications.

## Performance Errors and Patterns of Individual Differences

Theorists who argue that discrepancies between actual responses and those dictated by normative models are not indicative of human irrationality (e.g., Cohen, 1981) sometimes attribute the discrepancies to performance errors (see Stein, 1996, pp. 8–9). Borrowing the idea of a competence–performance distinction from linguists, these theorists view performance errors as the failure to apply a rule, strategy, or algorithm that is part of a person's competence because of a momentary and fairly random lapse in ancillary processes necessary to execute the strategy (e.g., lack of attention, temporary memory deactivation, and distraction).

This notion of a performance error as a momentary attention, memory, or processing lapse that causes responses to appear nonnormative, even when competence is fully normative, has implications for patterns of individual differences across reasoning tasks. For example, the strongest possible form of this view is that *all* discrepancies from normative responses are due to performance errors. This strong form has the implication that there should be virtually no correlations among performance on disparate reasoning tasks. If each departure from normative responding represents a momentary processing lapse that is due to distraction, carelessness, or temporary confusion, then there is no reason to expect covariance in performance across various indices of rational thinking. In contrast, positive manifold among disparate rational thinking tasks would call into question the notion that all variability in responding can be attributable to performance errors.

We have found very little evidence for the performance error view. With virtually all of the tasks from the heuristics and biases literature that we have examined, there is considerable internal consistency. Furthermore, at least for certain classes of task, there are significant cross-task correlations. The direction of these correlations is almost always the same — participants giving the normative response on one task are

TABLE 17.1 Intercorrelations Among Several Reasoning Tasks

| Variable | 1 | 2 | 3 |
|---|---|---|---|
| 1. Syllogisms | — | | |
| 2. Selection task | .363*** | — | |
| 3. Statistical reasoning | .334*** | .258*** | — |
| 4. Argument evaluation | .340*** | .310*** | .117 |

*Note.* $ns$ = 188–195.
***$p$ < .001, two-tailed.

usually significantly more likely to give it on another. A typical set of results (see Stanovich & West, 1998) is displayed in Table 17.1. The tasks shown here included a syllogistic reasoning task in which the believability of the conclusion contradicted logical validity. Next were five abstract selection task problems (Newstead & Evans, 1995; Wason, 1966). The third task was derived from the literature on statistical reasoning and was inspired by the work of Nisbett and Ross (1980). The fourth task was an argument evaluation task (Stanovich & West, 1997) that taps reasoning skills of the type studied in the informal reasoning literature (Baron, 1995; Klaczynski & Gordon, 1996; Klaczynski, Gordon, & Fauth, 1997; Kuhn, 1993).

The correlations among the four rational thinking tasks are displayed in Table 17.1. Five of the six correlations were significant at the .001 level. The significant relationships among most of the rational thinking tasks (which derive from very different reasoning domains) suggest that departures from normative responding on each of them were due to systematic factors and not to nonsystematic performance errors. On an individual task basis, however, most of the correlations were of a modest magnitude. Nevertheless, it should also be emphasized that many of the relationships might be underestimated due to modest reliability. Due to the logistical constraints of a multivariate investigation involving so many different tasks, scores on some of these measures were based on a few number of trials.

## Implications of Individual Differences for Prescriptive Models: Algorithmic-Level Limitations

Patterns of individual differences might have implications that extend beyond testing the view that discrepancies between descriptive models and normative models arise entirely from performance errors. Additionally, judgments about the rationality of actions and beliefs must take into account the resource-limited nature of the human cognitive apparatus (Baron, 1985; Cherniak, 1986; Goldman, 1978; Harman, 1995; Oaksford & Chater, 1993, 1995; Stich, 1990). The idea of computational limitations is best discussed by first making a distinction, popular in cognitive theory, between the algorithmic level of analysis (concerning the computational processes necessary to carry out a task) and the rational level of analysis that encompasses the goals of the system, the beliefs relevant to those goals, and the choice of action that is rational, given the system's goals and beliefs (see Anderson, 1990; Marr, 1982; Newell, 1982, 1990).

The important point for the present discussion is that even if all humans were optimally adapted to their environments at the rational level of analysis, there may still be computational limitations at the algorithmic level that prevent the normative response. Thus, the magnitude of the correlation between performance on a reasoning task and cognitive capacity provides an empirical clue about the importance of algorithmic limitations in creating discrepancies between descriptive and normative models. A strong correlation suggests important algorithmic limitations that might make the normative response not prescriptive for those of lower cognitive capacity. In contrast, the absence of a correlation between the normative response and cognitive capacity suggests no computational limitation and thus no reason why the normative response should not be considered prescriptive (see Baron, 1985; Bell, Raiffa, & Tversky, 1988).

In our studies, we have operationalized cognitive capacity in terms of well-known cognitive ability and academic aptitude tasks such as the Scholastic Aptitude Test (SAT). All are known to load highly on psychometric *g* (Carpenter, Just, & Shell, 1990; Carroll, 1993; Matarazzo, 1972), and such measures have been linked to neurophysiological and information-processing indicators of efficient cognitive computation (Deary & Stough, 1996; Detterman, 1994; Hunt, 1987; Stankov & Dunn, 1993; Vernon, 1991, 1993).

The top half of Table 17.2 indicates the magnitude of the correlation between SAT total scores and the four reasoning tasks discussed previously. SAT scores were significantly correlated with performance on all four rational thinking tasks. The

TABLE 17.2 Correlations Between the Reasoning Tasks and SAT Total Score

| Task | Correlation |
|---|---|
| Syllogisms | .470*** |
| Selection task | .394*** |
| Statistical reasoning | .347*** |
| Argument evaluation | .358*** |
| Replication and extension | |
| Syllogisms | .410*** |
| Statistical reasoning | .376*** |
| Argument evaluation task | .371*** |
| Covariation detection | .239*** |
| Hypothesis testing bias | −.223*** |
| Outcome bias | −.172*** |
| If/only thinking | −.208*** |
| RT1 composite | .530*** |
| RT2 composite | .383*** |
| RT composite (all tasks) | .547*** |

*Note.* For replication and extension, sample size ranged from 527 to 529. SAT = Scholastic Aptitude Test; RT1 composite = standard score composite of performance on argument evaluation task, syllogisms, and statistical reasoning; RT2 composite = standard score composite of performance on covariation judgment, hypothesis testing task, if/only thinking, and outcome bias; RT composite (all tasks) = rational thinking composite score of performance on all seven tasks in the replication and extension experiment.
***$p < .001$, two-tailed.

correlation with syllogistic reasoning was the highest (.470), and the other three correlations were roughly equal in magnitude (.347 to .394). All were statistically significant ($p < .001$).

The remaining correlations in the table are the results from a replication and extension experiment (see Stanovich & West, 1998). Three of the four tasks from the previous experiment were carried over (all but the selection task), and added to this multivariate battery was a covariation detection task. Three new tasks assessing cognitive biases were also added to this multivariate battery of tests. The first was a hypothesis testing task modeled on Tschirgi (1980) in which the score on the task was the number of times participants attempted to test a hypothesis in a manner that did not unconfound variables. Outcome biases was measured by using tasks introduced by Baron and Hershey (1988). This bias is demonstrated when participants rate a decision with a positive outcome 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 bias* refers to the tendency for people to have differential responses to outcomes on the basis of the differences in counterfactual alternative outcomes that might have occurred (Epstein, Lipson, Holstein, & Huh, 1992). The bias is demonstrated when participants 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.

The bottom half of Table 17.2 indicates that the correlations involving the syllogistic reasoning task, statistical reasoning task, and argument evaluation task were similar in magnitude to those obtained in the previous experiment. 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 nonnormative cognitive biases. 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 the table 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 rational thinking composite was formed by summing the standard scores of the remaining four tasks: covariation judgment, hypothesis testing, if/only thinking, outcome bias (the latter three scores are 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 rational thinking composites were combined into a composite variable reflecting performance on all seven tasks, and this composite displayed a correlation of .547 with SAT scores. It thus appears that, to a considerable extent, discrepancies between actual performance and normative models can be accounted for by variation in capacity limitations at the algorithmic level — at least with respect to the tasks investigated in this experiment. However, in the following experiments presented we examine individual differences in situations in which the interpretation of the gap between the descriptive and the normative is much more contentious.

## Applying the Right Normative Model

In addition to performance errors and algorithmic limitations, there are further reasons why observed performance might depart from normative prescriptions. For example, psychologists have traditionally appealed to the normative models of other disciplines (e.g., statistics, logic, mathematics, and decision science) to interpret the responses on various tasks. There is a danger in this procedure. The danger arises because there is a lack of consensus on the status of the normative models in many of the disciplines from which psychologists borrow. Heavy reliance on a normative model that is in dispute often engenders the claim that the gap between the descriptive and normative occurs because the psychologist is applying the wrong normative model to the situation; in short, the problem is with the experimenter and not with the participant (see Cosmides & Tooby, 1996; Levi, 1983; Lopes, 1981; Macdonald, 1986). For example, Birnbaum (1983) demonstrated that conceptualizing the well-known taxicab base-rate problem (see Bar-Hillel, 1980; Tversky & Kahneman, 1982) within a signal-detection framework can lead to different, normatively correct conclusions than those assumed by the less flexible Bayesian model that is usually applied. Likewise, Dawes (1989, 1990) and Hoch (1987) argued that social psychologists have too hastily applied an overly simplified normative model in labeling performance in opinion prediction experiments as displaying a so-called false consensus (see also Krueger & Clement, 1994; Krueger & Zeiger, 1993).

One way to test indirectly the claim that the wrong normative model is being applied is to investigate how responses on reasoning tasks correlate with measures of cognitive capacity. If that correlation is positive, it would seem to justify the use of the normative model being used to evaluate performance; whereas negative correlations might indicate that an inappropriate normative model is being applied to the situation. This would seem to follow from the arguments of the optimization theorists, who emphasize the adaptiveness of human cognition (Anderson, 1990, 1991; Cosmides & Tooby, 1994, 1996; Oaksford & Chater, 1993, 1994, 1995; Payne, Bettman, & Johnson, 1993; Schoemaker, 1991; Shanks, 1995). The responses of organisms with fewer algorithmic limitations would be assumed to be closer to the response that a rational analysis (Anderson, 1991) would reveal as optimal. For example, the optimal strategy might be computationally more complex, and only those with the requisite computational power might be able to compute it. Under standard assumptions about the adaptive allocation of cognitive resources (Anderson, 1991; Payne et al., 1993; Schoemaker, 1991), the additional computational complexity would only be worth dealing with if the strategy were indeed more efficacious. Alternatively, the optimal strategy might not be more computationally complex. It might simply be more efficient and more readily recognized as such by more intelligent organisms. Thus, negative correlations with the response considered normative might call into question the appropriateness of the normative model being applied.

With these arguments in mind, it is thus interesting to note that the direction of all of the correlations displayed in Table 17.2 (as well as in Table 17.1) is consistent with the standard normative models used by psychologists when interpreting tasks in the reasoning and decision-making literature. This is not always the case, however. We examine here a case of a task in which the normative model to be applied has been the subject of enormous dispute and in which our analysis of individual differences suggests that the model traditionally applied in the heuristics and biases literature may be questionable.

The statistical reasoning problems utilized in the experiments discussed so far (those derived from Fong, Krantz, & Nisbett, 1986) have a less controversial history because they involve causal aggregate information, analogous to the causal base rates discussed by Ajzen (1977) and Bar-Hillel (1980, 1990); that is, base rates that had a causal relationship to the criterion behavior. In contrast, noncausal base rates — those bearing no obvious causal relationship to the criterion behavior — have been the subject of over a decade's worth of contentious dispute (Bar-Hillel, 1990; Cohen, 1981, 1986; Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Kahneman & Tversky, 1996; Koehler, 1996; Levi, 1983). In several experiments (see Stanovich & West, 1998) we have included some of these noncausal base-rate problems that are notorious for provoking philosophical dispute. One was an AIDS testing problem modeled on Casscells, Schoenberger, and Graboys (1978):

> Imagine that AIDS occurs in one in every 1,000 people. Imagine also there is a test to diagnose the disease that always gives a positive result when a person has AIDS. Finally, imagine that the test has a false positive rate of 5 percent. This means that the test wrongly indicates that AIDS is present in 5 percent of the cases where the person does not have AIDS. Imagine that we choose a person randomly, administer the test, and that it yields a positive result (indicates that the person has AIDS). What is the probability that the individual actually has AIDS, assuming that we know nothing else about the individual's personal or medical history?

The Bayesian posterior probability for this problem is slightly less than .02. Thus, responses of less than 10% were interpreted as indicating Bayesian amalgamation, responses of over 90% were scored as indicating strong reliance on indicant information, and responses between 10% and 90% were scored as intermediate. Using this classification scheme, we classified 107 participants as strongly reliant on indicant information, 50 as intermediate, and 40 as approximately Bayesian.

As indicated in Table 17.3, the three groups displayed a significant difference in their mean total SAT scores. The mean SAT score of the participants strongly reliant on indicant information (1,115) was higher than the mean score of the Bayesian participants (1,071) whose mean was higher than that of the group showing moderate reliance on indicant information (1,061). Significant differences were also observed on the Raven Progressive Matrices Test (Raven, 1962), Nelson–Denny Comprehension Test (Brown, Bennett, & Hanna, 1981), and a syllogistic reasoning task. In each case, the indicant participants outperformed the other two groups. No significant differences were obtained on the selection task, statistical reasoning task, and argument evaluation task, although the differences tended in the same direction.

Exactly the same trends were apparent in a replication experiment displayed at the bottom of Table 17.3. The mean SAT score of the participants strongly reliant on indicant information (1,153) was significantly higher than the mean score of either the Bayesian participants (1,103) or the intermediate participants (1,109). There was a statistically significant difference in the group mean scores on the composite score for the causal aggregate statistical reasoning problems. Most interesting, however, was the direction of the differences. The highest mean score was achieved by the group highly reliant on the indicant information in the AIDS problem, followed by the mean of the group showing moderate reliance on indicant

TABLE 17.3 Mean Task Performance for the Groups Classified as Indicant, Intermediate, and Bayesian on the AIDS Problem

| Task | Indicant ($n$ = 107) | Intermediate ($n$ = 50) | Bayesian ($n$ = 40) | *dfs* | *F* |
|---|---|---|---|---|---|
| SAT total | $1{,}115_a$ | $1{,}061_b$ | 1,071 | 2, 181 | 4.26 |
| Raven Matrices | $10.09_a$ | $8.56_b$ | 9.40 | 2, 194 | 4.82** |
| Nelson–Denny | 20.23 | 19.52 | 19.05 | 2, 194 | 3.09* |
| Syllogisms | $4.79_a$ | $3.66_b$ | 4.21 | 2, 192 | 4.65 |
| Selection task | 1.61 | 1.46 | 1.11 | 2, 188 | 0.48 |
| Statistical reasoning | 0.421 | −0.472 | −0.537 | 2, 194 | 2.42 |
| Argument evaluation | 0.345 | 0.322 | 0.351 | 2, 191 | 0.30 |
| | | Replication | | | |
| | ($n$ = 118) | ($n$ = 57) | ($n$ = 36) | | |
| SAT total | $1{,}153_a$ | $1{,}109_b$ | $1{,}103_b$ | 2, 198 | 4.60 |
| Raven Matrices | 9.49 | 9.04 | 8.64 | 2, 189 | 0.89 |
| Nelson–Denny | 20.47 | 20.08 | 19.47 | 2, 194 | 1.95 |
| Syllogisms | 5.40 | 4.93 | 4.92 | 2, 208 | 1.32 |
| Statistical reasoning | $0.726_a$ | $-0.840_b$ | $-1.051_b$ | 2, 208 | 7.24** |

*Note.* Means with different subscripts are significantly different (Scheffé). SAT = Scholastic Aptitude Test.
$^*p < .05$. $^{**}p < .01$.

information. The participants giving the Bayesian answer on the AIDS problem were *least* reliant on the aggregate information in the causal statistical reasoning problems.

The results from both of these experiments indicate that the noncausal base-rate problems display patterns of individual differences quite unlike those shown on the causal aggregate problems. On the latter, participants giving the statistical response (choosing the aggregate rather than the case or indicant information) scored consistently higher on measures of cognitive ability and were disproportionately likely to give the standard normative response on other rational thinking tasks (see Tables 17.1 and 17.2). This pattern did not hold for the AIDS problem in which the significant differences were in the opposite direction: participants strongly reliant on the indicant information scored higher on measures of cognitive ability and were more likely to give the standard normative response on other rational thinking tasks, including other base-rate problems (of the causal variety, see Bar-Hillel, 1980, 1990). Interestingly, the AIDS problem (or close variants of it) has been the focus of intense debate in the literature, and several authors have argued against making the automatic assumption that the indicant response is nonnormative in the version that we had utilized.

## Which Task Construals Are Associated With Differences in Cognitive Capacity?

Theorists who resist attributing irrational cognition as a cause of the gap between normative and descriptive models have one more strategy in addition to those described previously. It is the argument that researchers may well be applying the correct normative model to the problem as set but that the participant might have construed the problem differently and be providing the normatively appropriate answer to a different problem (Adler, 1984, 1991; Hilton, 1995; Levinson, 1995; Margolis, 1987; Schwarz, 1996). Such an argument is somewhat different from any of the critiques that have been mentioned so far. It is not the equivalent of positing that a performance error has been made, because performance errors (e.g., attention lapses and temporary memory lapse) would not be expected to recur in exactly the same way in a readministration of the same task. In contrast, if the participant has truly misunderstood the task, he or she would be expected to do so again on an identical readministration of the task.

Correspondingly, this criticism is quite different from the argument that the task exceeds the computational capacity of the participant. The latter explanation puts the onus of the suboptimal performance on the participant. In contrast, the alternative task construal argument places the blame at least somewhat on the shoulders of the experimenter for failing to realize that there were task features that might lead participants to frame the problem in a manner different from that intended. In locating the problem with the experimenter, it is similar to the wrong norm explanation. However, it is different in that in the latter, it is assumed that the participant 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 is that the experimenter may be applying the correct normative model to the problem the experimenter intends the participant to solve, but the participant might have construed the problem in some other way and be providing a normatively appropriate answer to a *different* problem.

An example of the alternative task construal interpretation is provided by one of the most famous problems in the heuristics and biases literature, the so-called Linda Problem (Tversky & Kahneman, 1983):

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
>
> a. Linda is a teacher in an elementary school
> b. Linda works in a bookstore and takes Yoga classes
> c. Linda is active in the feminist movement
> d. Linda is a psychiatric social worker
> e. Linda is a member of the League of Women Voters
> f. Linda is a bank teller
> g. Linda is an insurance salesperson
> h. Linda is a bank teller and is active in the feminist movement

Because Alternative h is the conjunction of Alternatives c and f, the probability of h cannot be higher than that of either Alternative c or Alternative f, yet 85% of the participants in Tversky and Kahneman's (1983) study rated Alternative h as more probable than Alternative f. What concerns us here is the argument that there are subtle linguistic and pragmatic features of the problem that lead the participant to evaluate alternatives different than those listed. For example, Hilton (1995) argued that under the assumption that the detailed information given about the target means that the experimenter knows a considerable amount about Linda, then it is reasonable to think that the phrase "Linda is a bank teller" does not contain the phrase "and is not active in the feminist movement" because the experimenter already knows this to be the case. If "Linda is a bank teller" is interpreted in this way, then rating Alternative h as more probable than Alternative f no longer represents a conjunction fallacy. Several other investigators have suggested that pragmatic inferences lead to seeming violations of the logic of probability theory in the Linda Problem (see Adler, 1991; Dulany & Hilton, 1991; Politzer & Noveck, 1991). These criticisms all share the implication that actually displaying the conjunction fallacy is a rational response to an alternative construal of the different statements to be ranked.

In a recent study (Stanovich & West, in press-b), we have examined the question of whether the different construals of the task are associated with differences in cognitive ability. That is, assuming that those displaying the so-called conjunction fallacy are making a nonextensional interpretation and that those avoiding the so-called fallacy are making the extensional interpretation that the investigators intended, we asked whether the participants making the nonextensional, pragmatic interpretation were participants who were disproportionately the participants of higher cognitive ability. Because this group is in fact the majority in most studies — and because the use of such pragmatic cues and background knowledge is often interpreted as reflecting adaptive information processing (e.g., Hilton, 1995) — optimization models of human cognition (e.g., Anderson, 1990) might be thought to predict that they would be the participants of higher computational capacity.

In our study, we examined the performance of 150 participants on the Linda Problem presented previously. Consistent with the results of previous experiments on this problem (Tversky & Kahneman, 1983), 80.7% of our sample (121 participants) displayed the conjunction effect — they rated the feminist bank teller alternative as more probable than the bank teller alternative. However, the individuals who displayed the conjunction effect had mean SAT scores 82 points below the mean of the individuals who did not display the effect (1,080 vs. 1,162), $t(148) = 3.58$, $p < .001$. This difference is sizable — translating into an effect size of .746, which Rosenthal and Rosnow (1991, p. 446) classified as "large."

Another problem that has spawned many arguments about alternative construals is Wason's (1966) selection task, mentioned briefly earlier (see Stanovich & West, in press-a). The participant is shown four cards lying on a table showing two letters and two numbers (A, D, 3, 7). They are told that each card has a number on one side and a letter on the other and that the experimenter has the following rule (of the if *P*, then *Q* type) in mind with respect to the four cards: "If there is an A on one side, then there is a 3 on the other." The participant is then told that he or she must turn over whichever cards are necessary to determine whether the experimenter's rule is true or false.

Performance on such abstract versions of the selection task is extremely low. Typically, less than 10% of participants make the correct selections of the A card (*P*) and 7 card (not *Q*). The most common incorrect choices made by participants are the A card and the 3 card (*P* and *Q*) or the selection of the A card only (*P*). The preponderance of *P* and *Q* responses has most often been attributed to a so-called matching bias that is automatically triggered by surface-level relevance cues (Evans, 1996; Evans, Newstead, & Byrne, 1993), but some investigators have championed an explanation that is based on an alternative task construal. For example, Oaksford and Chater (1994, 1996; see also Nickerson, 1996) argued that rather than interpreting the task as one of deductive reasoning (as the experimenter intends), many participants interpret it as an inductive problem of probabilistic hypothesis testing. They showed that the *P* and *Q* response is then dictated under a formal Bayesian analysis that assumes such an interpretation.

Table 17.4 presents the mean SAT scores of participants giving a variety of response combinations to a selection task problem (see Stanovich & West, in press-a). Respondents giving the deductively correct *P* and not-*Q* response had the highest SAT scores followed by the participants choosing the *P* card only. All other responses, including the modal *P* and *Q* response, were given by participants having SAT scores some 100 points lower than those giving the correct response under a deductive construal.

One possible interpretation of the individual differences displayed on the Linda Problem and on the selection task is in terms of two-process theories of reasoning (Epstein, 1994; Evans, 1984, 1996; Evans & Over, 1996; Sloman, 1996). For example, Sloman distinguished an associative processing system with computational mechanisms that reflect similarity and temporal contiguity and a rule-based system that operates on symbolic structures having logical content. According to Sloman, the associative system responds to the similarity in the Linda Problem ("representativeness," in the terminology of Tversky & Kahneman, 1983); whereas the rule-based system engages extensional probabilistic concepts that dictate that the bank teller alternative is more probable.

We conjecture here that large differences in cognitive ability will be found only in problems that strongly engage both reasoning systems and that cue opposite responses. This is because the two systems are identified with different types of intelligence. Clearly, the rule-based system embodies analytic intelligence of the type measured on SAT tests (Carpenter et al., 1990; Carroll, 1993). The associative system, in contrast, might be better identified with what Levinson (1995; see also Cummins, 1996) termed *interactional intelligence*. He speculated that evolutionary pressures were focused more on negotiating cooperative mutual intersubjectivity than on understanding the natural world. Having as its goals the ability to model other minds to read intention and to make rapid interactional moves on the basis of those modeled intentions, interactional intelligence is composed of pragmatic heuristics that operate to facilitate intention attribution.

**TABLE 17.4** Mean SAT Total Scores as a Function of Response Given on the Selection Task

| Response | Score | *n* |
|---|---|---|
| Correct | 1,190 | 24 |
| *P* | 1,150 | 38 |
| All | 1,101 | 21 |
| *P*, *Q* | 1,095 | 144 |
| *P*, *Q*, *NQ* | 1,084 | 14 |
| Other | 1,070 | 53 |

*Note.* SAT = Scholastic Aptitude Test.

If the two systems cue opposite responses in a particular task, the rule-based system will tend to cue differentially those of high analytic intelligence, and this tendency will not be diluted by the associative system nondifferentially drawing participants to the same response (because the associative system is unrelated to analytic intelligence; see Reber, 1993). The Linda Problem maximizes the tendency for the associative and rule-based systems to prime different responses, and this problem displayed a large difference in cognitive ability. The selection task might likewise maximize the tendency for the associative and rule-based systems to prime different responses.

## Thinking Dispositions and Individual Differences in Rational Thought

In several studies in our research program (see Stanovich & West, 1997, 1998), we have examined one other critical issue — whether there is reliable variance in performance on rational thinking tasks after differences in computational power have been accounted for and whether this residual variation is associated with cognitive strategies, styles, propensities, or dispositions. The conceptual basis for this aspect of our research resides in models of thinking that distinguish between cognitive capacities and thinking dispositions (e.g., Baron, 1985, 1994; Klaczynski et al., 1997; Norris, 1992; Sternberg & Ruzgis, 1994). For example, it is possible that these two constructs (cognitive ability and thinking dispositions) are actually at different levels of analysis in a cognitive theory and that they do separate explanatory work. Variation in cognitive ability refers to individual differences in the efficiency of processing at the algorithmic level. In contrast, thinking dispositions of the type studied in this investigation elucidate individual differences at the rational level. They are telling us about the individual's goals and epistemic values (see Kruglanski, 1989; Kruglanski & Webster, 1996).

With regard to thinking dispositions, we focused on those most relevant to epistemic rationality — processes leading to more accurate belief formation and to more consistent belief–desire networks (Harman, 1995; Stanovich, 1994, in press; Stanovich & West, 1997; Thagard, 1992). We attempted to tap the following dimensions: epistemological absolutism, willingness to perspective switch, willingness to decontextualize, and tendency to consider alternative opinions and evidence.

In the case of many of the individual tasks examined in our research program, thinking dispositions of this type do in fact predict residual variance. In lieu of displaying all of these analyses (see Stanovich & West, 1997, 1998), we present an analysis that is based on composite variables. The variance partitioning is displayed in Table 17.5 in the form of a commonality analysis. The criterion variable was the first rational thinking composite score discussed earlier, reflecting combined performance on the argument evaluation, statistical reasoning, and syllogistic reasoning tasks. SAT total scores and the thinking dispositions composite score attained a multiple $R$ with this criterion variable of .600, $F(2, 526) = 148.15$, $p < .001$. Thus, a substantial amount of variance (36%) on these rational thinking tasks is jointly explained by these two predictors. SAT total was a significant unique predictor (partial correlation = .478, unique variance explained = .190), $F(1, 526) = 156.17$, $p < .001$, as was the thinking dispositions composite score (partial correlation =

**TABLE 17.5** Commonality Analysis on Rational Thinking Composite Score

| Variable | Total: Unique | Total: Common | Variance explained |
|---|---|---|---|
| SAT | .190*** | .091*** | |
| TDC composite | .079*** | .091*** | |
| | | | .360 |

*Note.* SAT = Scholastic Aptitude Test; TDC = thinking dispositions composite.
***$p < .001$, two-tailed.

.332, unique variance explained = .079), $F(1, 526) = 65.03$, $p < .001$. Correlations and partial correlations involving these variables are presented in Table 17.6.

In several other analyses structurally similar to this one, dispositions toward open-minded and counterfactual thinking, and the lack of dogmatic and absolutist thinking, were associated with superior performance on rational thinking tasks, even after the variance accounted for by several measures of general cognitive ability had been partialled out. These results support the distinction between thinking dispositions and cognitive capacities that is championed by some investigators (e.g., Baron, 1985) and validate the increasing attention that is being given to processes that are at the borderline of cognitive psychology and personality research (Ackerman & Heggestad, 1997; Goff & Ackerman, 1992; Klaczynski & Gordon, 1996; Sternberg, 1997; Sternberg & Ruzgis, 1994). Thinking dispositions of the type we have examined may provide information about epistemic goals at the rational level of analysis (see Anderson, 1990). What such a result may be telling researchers is that to understand variation in reasoning in such tasks, they need to examine more than just differences at the algorithmic level (computational capacity). In addition, the epistemic goals of the reasoners must also be examined.

The importance of thinking styles in discussions of human rationality has perhaps not received sufficient attention because of the heavy reliance on the competence–performance distinction in philosophical treatments of rational thought in which all of the important psychological mechanisms are allocated to the competence side of the dichotomy. Cohen (1982), for example, argued that there are really only two

**TABLE 17.6** Correlations and Partial Correlations

| Variable | 1 | 2 | 3 |
|---|---|---|---|
| 1. SAT | | .53*** | .27*** |
| 2. RT1 composite | .48*** | | .41*** |
| 3. TDC composite | .06 | .33*** | |

*Note.* Zero-order correlations are shown above the diagonal, and partial correlations are shown below the diagonal. SAT = Scholastic Aptitude Test; RT1 composite = standard score composite of performance on argument evaluation task, syllogisms, and statistical reasoning; TDC = thinking dispositions composite score.
***p < .001, two-tailed.

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 surprising, given such a conceptualization, the processes contributing to error ("adventitious causes") are of little interest to Cohen (1981, 1982). There is nothing in such a view that would motivate any interest in patterns of errors or individual differences in such errors.

In contrast, Johnson-Laird and Byrne (1993) articulated a view of rational thought that parses the competence–performance distinction much differently from that of Cohen (1981, 1982, 1986) and that simultaneously leaves room for cognitive styles to play an important role in determining responses when people face situations in which problem solving or decision making is required. At the heart of the rational competence that Johnson-Laird and Byrne (1993) attributed to humans is only one metaprinciple: People are programmed to accept inferences as valid provided that they have constructed no mental model of the premises that contradict the inference. Inferences are categorized as false when a mental model is discovered that is contradictory. However, the search for contradictory models is "not governed by any systematic or comprehensive principles" (Johnson-Laird & Byrne, 1993, p. 178). In this passage, Johnson-Laird and Byrne seem to be arguing that there are no systematic control features of the search process. However, epistemically related cognitive dispositions may in fact be reflecting just such control features. Individual differences in the extensiveness of the search for contradictory models could arise from a variety of cognitive factors that, although they may not be completely systematic, may be far from adventitious — factors such as dispositions toward premature closure, cognitive confidence, reflectivity, dispositions toward confirmation bias, ideational generativity, and so forth.

## Conclusions

In our research program, we have attempted to demonstrate that a consideration of individual differences in the heuristics and biases literature may have implications for debates about theories of the gap between normative models and descriptive models of actual performance. In reply to Cohen's (1981) well-known critique of the heuristics and biases 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 largely been ignored in more recent debates about human rationality and the tasks that we use to assess it (for exceptions, see Slugoski, Shields, & Dawson, 1993; Stankov & Crawford, 1996; Yates, Lee, & Shinotsuka, 1996). For example, philosopher Nicholas Rescher (1988) argued that

> to construe the data of these interesting experimental studies [of probabilistic reasoning] 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 focus — like that of many who have dealt with the philosophical implications of the idea of human irrationality — is on the issue of how humans are systematically programmed. Inclinations toward questionable suppositions are of interest only 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 (Baron, 1994; Piattelli-Palmarini, 1994; Shafir, 1994; Shafir & Tversky, 1995; Wagenaar, 1988).

One of the purposes of the present research program was to reverse the figure and ground in this dispute, which has tended to be dominated by the particular way that philosophers frame the competence–performance distinction. Specifically, from a psychological standpoint, there may be important implications in precisely the aspects of performance that have been backgrounded in the controversy about basic reasoning competence. That is, whatever the outcome of the disputes about how humans are "systematically programmed" (Cosmides, 1989; Johnson-Laird & Byrne, 1991, 1993; Oaksford & Chater, 1993, 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 reported here indicate that, at least for certain subsets of tasks, the "inclination toward questionable suppositions" has some degree of domain generality, it is in some cases linked to computational limitations, and it is sometimes predicted by thinking dispositions that can be related to the epistemic and pragmatic goals of rational thought.

## REFERENCES

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219–245.

Adler, J. E. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behaviour, 14*, 165–181.

Adler, J. E. (1991). An optimist's pessimism: Conversation and conjunctions. In E. Eells & T. Maruszewski (Eds.), *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science* (pp. 251–282). Amsterdam: Editions Rodopi.

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology, 35*, 303–314.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14*, 471–517.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44*, 211–233.

Bar-Hillel, M. (1990). Back to base rates. In R. M. Hogarth (Ed.), *Insights into decision making: A tribute to Hillel J. Einhorn* (pp. 200–216). Chicago: University of Chicago Press.

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

Baron, J. (1994). *Thinking and deciding* (2nd ed.). Cambridge, England: Cambridge University Press.

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

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*, 569–579.

Bell, D., Raiffa, H., & Tversky, A. (Eds.). (1988). *Decision making: Descriptive, normative, and prescriptive interactions.* Cambridge, England: Cambridge University Press.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96*, 85–94.

Brown, J., Bennett, J., & Hanna, G. (1981). *The Nelson–Denny Reading Test.* Lombard, IL: Riverside Publishing Co.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.

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

Cherniak, C. (1986). *Minimal rationality.* Cambridge, MA: MIT Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317–370.

Cohen, L. J. (1982). Are people programmed to commit fallacies? *Journal for the Theory of Social Behavior, 12*, 251–274.

Cohen, L. J. (1986). *The dialogue of reason.* London: Oxford University Press.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Watson selection task. *Cognition, 31*, 187–276.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition, 50*, 41–77.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1–73.

Cummins, D. D. (1996). Evidence for innateness of deontic reasoning. *Mind and Language, 11*, 160–190.

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology, 25*, 1–17.

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

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time. *American Psychology, 51*, 599–608.

Detterman, D. K. (1994). Intelligence and the brain. In P. A. Vernon (Ed.), *The neuropsychology of individual differences* (pp. 35–57). San Diego, CA: Academic Press.

Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, 9, 85–110.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709–724.

Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology, 62*, 328–339.

Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 451–468.

Evans, J. St. B. T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology, 87*, 223–240.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction.* Hove, England: Erlbaum.

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

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*, 253–292.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103*, 592–596.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.

Goff, M., & Ackerman, P. L. (1992). Personality–intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology, 84*, 537–552.

Goldman, A. I. (1978). Epistemics: The regulative theory of cognition. *Journal of Philosophy, 55*, 509–523.

Harman, G. (1995). Rationality. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol. 3, pp. 175–211). Cambridge, MA: MIT Press.

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

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology, 53,* 221–234.

Hunt, E. (1987). The next word on verbal ability. In P. A. Vernon (Ed.), *Speed of information-processing and intelligence* (pp. 347–392). Norwood, NJ: Ablex.

Jepson, C., Krantz, D., & Nisbett, R. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences, 6,* 494–501.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (1993). Models and deductive rationality. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 177–210). London: Routledge & Kegan Paul.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103,* 582–591.

Klaczynski, P. A., & Gordon, D. H. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology, 62,* 317–339.

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

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences, 19,* 1–53.

Krueger, J., & Clement, R. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 65,* 596–610.

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

Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases.* New York: Plenum Press.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing the mind: "Seizing" and "freezing." *Psychological Review, 103,* 263–283.

Kuhn, D. (1993). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly, 38,* 74–103.

Levi, I. (1983). Who commits the base rate fallacy? *Behavioral and Brain Sciences, 6,* 502–506.

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

Lopes, L. L. (1981). Performing competently. *Behavioral and Brain Sciences, 4,* 343–344.

Macdonald, R. (1986). Credible conceptions and implausible probabilities. *British Journal of Mathematical and Statistical Psychology, 39,* 15–27.

Margolis, H. (1987). *Patterns, thinking, and cognition.* Chicago: University of Chicago Press.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

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

Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18,* 87–127.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newstead, S. E., & Evans, J. St. B. T. (Eds.). (1995). *Perspectives on thinking and reasoning.* Hove, England: Erlbaum.

Nickerson, R. S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning, 2,* 1–31.

Nisbett, L., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Norris, S. P. (1992). Testing for the disposition to think critically. *Informal Logic, 14,* 157–164.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 31–60). London: Routledge & Kegan Paul.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631.

Oaksford, M., & Chater, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning, 1,* 121–152.

Osherson, D. N. (1995). Probability judgment. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol. 3, pp. 35–75). Cambridge, MA: MIT Press.

Payne, J. W., Bettman, J. R., & Johnson, E. (1993). *The adaptive decision maker.* Cambridge, England: Cambridge University Press.
Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds.* New York: Wiley.
Politzer, G., & Noveck, I. A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research, 20,* 83–103.
Raven, J. C. (1962). *Advanced progressive matrices (Set II).* London: H. K. Lewis.
Reber, A. S. (1993). *Implicit learning and tacit knowledge.* New York: Oxford University Press.
Rescher, N. (1988). *Rationality: A philosophical inquiry into the nature and rationale of reason.* New York: Oxford University Press.
Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.
Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
Schoemaker, P. (1991). The quest for optimality: A positive heuristic of science? *Behavioral and Brain Sciences, 14,* 205–245.
Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation.* Mahweh, NJ: Erlbaum.
Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition, 50,* 403–430.
Shafir, E., & Tversky, A. (1995). Decision making. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol. 3, pp. 77–100). Cambridge, MA: MIT Press.
Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology, 48A,* 257–279.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Review, 119,* 3–22.
Slugoski, B. R., Shields, H. A., & Dawson, K. A. (1993). Relation of conditional reasoning to heuristic processing. *Personality and Social Psychology Bulletin, 19,* 158–166.
Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences, 21,* 971–986.
Stankov, L., & Dunn, S. (1993). Physical substrata of mental energy: Brain capacity and efficiency of cerebral metabolism. *Learning and Individual Differences, 5,* 241–257.
Stanovich, K. E. (1994). Reconceptualizing intelligence: Dysrationalia as an intuition pump. *Educational Researcher, 23*(4), 11–22.
Stanovich, K. E. (in press). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.
Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89,* 342–357.
Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127,* 161–188.
Stanovich, K. E., & West, R. F. (in press-a). Cognitive ability and variation in selection task performance. *Thinking and Reasoning.*
Stanovich, K. E., & West, R. F. (in press-b). Individual differences in framing and conjunction effects. *Thinking and Reasoning.*
Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science.* London: Oxford University Press.
Sternberg, R. J. (1997). *Thinking styles.* Cambridge, England: Cambridge University Press.
Sternberg, R. J., & Ruzgis, P. (Eds.). (1994). *Personality and intelligence.* Cambridge, England: Cambridge University Press.
Stich, S. (1990). *The fragmentation of reason.* Cambridge, MA: MIT Press.
Thagard, P. (1992). *Conceptual revolutions.* Princeton, NJ: Princeton University Press.
Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51,* 1–10.
Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge, England: Cambridge University Press.
Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Vernon, P. A. (1991). The use of biological measures to estimate behavioral intelligence. *Educational Psychologist, 25*, 293–304.

Vernon, P. A. (1993). *Biological approaches to the study of human intelligence*. Norwood, NJ: Ablex.

Wagenaar, W. A. (1988). *Paradoxes of gambling behavior*. Hove, England: Erlbaum.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmonsworth, England: Penguin.

Yates, J. F., Lee, J., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes, 65*, 138–147.