BEHAVIORAL AND BRAIN SCIENCES (2000) **23,** 645–726
Printed in the United States of America

Individual differences in reasoning: Implications for the rationality debate?

Keith E. Stanovich
Department of Human Development and Applied Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1V6
kstanovich@oise.utoronto.ca

Richard F. West
School of Psychology, James Madison University, Harrisonburg, VA 22807
westrf@jmu.edu **falcon.jmu.edu//~westrf**

Abstract: Much research in the last two decades has demonstrated that human responses deviate from the performance deemed normative according to various models of decision making and rational judgment (e.g., the basic axioms of utility theory). This gap between the normative and the descriptive can be interpreted as indicating systematic irrationalities in human cognition. However, four alternative interpretations preserve the assumption that human behavior and cognition is largely rational. These posit that the gap is due to (1) performance errors, (2) computational limitations, (3) the wrong norm being applied by the experimenter, and (4) a different construal of the task by the subject. In the debates about the viability of these alternative explanations, attention has been focused too narrowly on the modal response. In a series of experiments involving most of the classic tasks in the heuristics and biases literature, we have examined the implications of individual differences in performance for each of the four explanations of the normative/descriptive gap. Performance errors are a minor factor in the gap; computational limitations underlie non-normative responding on several tasks, particularly those that involve some type of cognitive decontextualization. Unexpected patterns of covariance can suggest when the wrong norm is being applied to a task or when an alternative construal of the task should be considered appropriate.

Keywords: biases; descriptive models; heuristics; individual differences; normative models; rationality; reasoning

1. Introduction

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, they display confirmation bias, they test hypotheses inefficiently, they violate the axioms of utility theory, they do not properly calibrate degrees of belief, they overproject their own opinions on 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 1994; 1998; Evans 1989; Evans & Over 1996; Kahneman et al. 1982; Newstead & Evans 1995; Nickerson 1998; Osherson 1995; Piattelli-Palmarini 1994; Plous 1993; Reyna et al., in press; Shafir 1994; 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 Arkes & Hammond 1986; Kahneman et al. 1982).

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 (a substantial portion of that debate appearing in this journal; for summaries, see Baron 1994; Cohen 1981; 1983; Evans & Over 1996; Gigerenzer 1996a; Kahneman 1981; Kahneman & Tversky 1983; 1996; Koehler 1996; Stein 1996). The de-

Keith E. Stanovich is Professor of Human Development and Applied Psychology at the University of Toronto. He is the author of more than 125 scientific articles in the areas of literacy and reasoning, including *Who Is Rational? Studies of Individual Differences in Reasoning* (Erlbaum, 1999). He is a Fellow of the APA and APS and has received the Sylvia Scribner Award from the American Educational Research Association for contributions to research.

Richard F. West is a Professor in the School of Psychology at James Madison University, where he has been named a Madison Scholar. He received his Ph.D. in Psychology from the University of Michigan. The author of more than 50 publications, his main scientific interests are the study of rational thought, reasoning, decision making, the cognitive consequences of literacy, and cognitive processes of reading.

bate 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. Owing to the emphasis that these theorists place on reforming human cognition, they were labeled the Meliorists by Stanovich (1999). Disputing this contention were numerous investigators (termed the Panglossians; see 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 ascription of irrationality to subjects. First, instances of 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). 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).

However, in referring to the various alternative explanations (other than systematic irrationality) for the normative/descriptive gap, Rips (1994) warns that "a determined skeptic can usually explain away any instance of what seems at first to be a logical mistake" (p. 393). In an earlier criticism of Henle's (1978) Panglossian position, Johnson-Laird (1983) made the same point: "There are no criteria independent of controversy by which to make a fair assessment of whether an error violates logic. It is not clear what would count as crucial evidence, since it is always possible to provide an alternative explanation for an error" (p. 26). The most humorous version of this argument was made by Kahneman (1981) in his dig at the Panglossians who seem to have only two categories of errors, "pardonable errors by subjects and unpardonable ones by psychologists" (p. 340). Referring to the four classes of alternative explanations discussed above – performance errors, computational limitations, alternative problem construal, and incorrect norm application – Kahneman notes that Panglossians have "a handy kit of defenses that may be used if [subjects are] accused of errors: temporary insanity, a difficult childhood, entrapment, or judicial mistakes – one of them will surely work, and will restore the presumption of rationality" (p. 340).

These comments by Rips (1994), Johnson-Laird (1983), and Kahneman (1981) highlight the need for principled constraints on the alternative explanations of normative/descriptive discrepancies. In this target article we describe a research logic aimed at inferring such constraints from patterns of individual differences that are revealed across a wide range of tasks in the heuristics and biases literature. We argue here – using selected examples of empirical results (Stanovich 1999; Stanovich & West 1998a; 1998b; 1998c; 1998d; 1999) – that these individual differences and their patterns of covariance have implications for explanations of why human behavior often departs from normative models.[1]

2. Performance errors

Panglossian 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. Borrowing the idea of a competence/performance distinction from linguists (see Stein 1996, pp. 8–9), 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 (lack of attention, temporary memory deactivation, distraction, etc.). Stein (1996) explains the idea of a performance error by referring to a "mere mistake" – a more colloquial notion that involves "a *momentary lapse,* a divergence from some typical behavior. This is in contrast to attributing a divergence from norm to reasoning in accordance with principles that diverge from the normative principles of reasoning. Behavior due to irrationality connotes a *systematic* divergence from the norm" (p. 8). Similarly, in the heuristics and biases literature, the term bias is reserved for systematic deviations from normative reasoning and does not refer to transitory processing errors ("a bias is a source of error which is systematic rather than random," Evans 1984, p. 462).

Another way to think of the performance error explanation is to conceive of it within the true score/measurement error framework of classical test theory. Mean or modal performance might be viewed as centered on the normative response – the response all people are trying to approximate. However, scores will vary around this central tendency due to random performance factors (error variance).

It should be noted that Cohen (1981) and Stein (1996) sometimes encompass computational limitations within their notion of a performance error. In the present target article, the two are distinguished even though both are identified with the algorithmic level of analysis (see Anderson 1990; Marr 1982; and the discussion below on levels of analysis in cognitive theory) because they have different implications for covariance relationships across tasks. Here, performance errors represent algorithmic-level problems that are transitory in nature. Nontransitory problems at the algorithmic level that would be expected to recur on a readministration of the task are termed computational limitations.

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 of the hypothesis has the implication that there should be virtually no correlations among nonnormative processing biases across tasks. If each departure from normative responding represents a momentary processing lapse due to distraction, carelessness, or temporary confusion, then there is no reason to expect covariance among biases across tasks (or covariance among items *within* tasks, for that matter) because error variances should be uncorrelated.

In contrast, positive manifold (uniformly positive bivariate associations in a correlation matrix) among disparate tasks in the heuristics and biases literature – and among items within tasks – would call into question the notion that all variability in responding can be attributable to performance errors. This was essentially Rips and Conrad's (1983) argument when they examined individual differences in deductive reasoning: "Subjects' absolute scores on the propositional tests correlated with their performance on certain other reasoning tests. . . . If the differences in propositional reasoning were merely due to interference from other performance factors, it would be difficult to explain why they correlate with these tests" (pp. 282–83). In fact, a parallel argument has been made in economics where, as in reasoning, models of perfect market rationality are protected from refutation by positing the existence of local market mistakes of a transitory nature (temporary information deficiency, insufficient attention due to small stakes, distractions leading to missed arbitrage opportunities, etc.).

Advocates of perfect market rationality in economics admit that people make errors but defend their model of idealized competence by claiming that the errors are essentially random. The following defense of the rationality assumption in economics is typical in the way it defines performance errors as unsystematic: "In mainstream economics, to say that people are rational is not to assume that they never make mistakes, as critics usually suppose. It is merely to say that they do not make systematic mistakes – i.e., that they do not keep making the same mistake over and over again" (*The Economist* 1998, p. 80). Not surprising, others have attempted to refute the view that the only mistakes in economic behavior are unpredictable performance errors by pointing to the systematic nature of some of the mistakes: "The problem is not just that we make random computational mistakes; rather it is that our judgmental errors are often systematic" (Frank 1990, p. 54). Likewise, Thaler (1992) argues that "a defense in the same spirit as Friedman's is to admit that of course people make mistakes, but the mistakes are not a problem in explaining aggregate behavior as long as they tend to cancel out. Unfortunately, this line of defense is also weak because many of the departures from rational choice that have been observed are systematic" (pp. 4–5). Thus, in parallel to our application of an individual differences methodology to the tasks in the heuristics and biases literature, Thaler argues that variance and covariance patterns can potentially falsify some applications of the performance error argument in the field of economics.

Thus, as in economics, we distinguish systematic from unsystematic deviations from normative models. The latter we label performance errors and view them as inoculating against attributions of irrationality. Just as random, unsystematic errors of economic behavior do not impeach the model of perfect market rationality, transitory and random errors in thinking on a heuristics and biases problem do not impeach the Panglossian assumption of ideal rational competence. Systematic and repeatable failures in algorithmic-level functioning likewise do not impeach intentional-level rationality, but they are classified as computational limitations in our taxonomy and are discussed in section 3. Systematic mistakes not due to algorithmic-level failure do call into question whether the intentional-level description of behavior is consistent with the Panglossian assumption of perfect rationality – provided the normative model being applied is not inappropriate (see sect. 4) or that the subject has not arrived at a different, intellectually defensible interpretation of the task (see sect. 5).

In several studies, we have found very little evidence for the strong version of 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. Further, at least for certain classes of task, there are significant cross-task correlations. For example, in two different studies (Stanovich & West 1998c) we found correlations in the range of .25 to .40 (considerably higher when corrected for attenuation) among the following measures:

1. Nondeontic versions of Wason's (1966) selection task: The subject 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 subject is then told that he/she must turn over whichever cards are necessary to determine whether the experimenter's rule is true or false. Only a small number of subjects make the correct selections of the A card (P) and 7 card (not-Q) and, as a result, the task has generated a substantial literature (Evans et al. 1993; Johnson-Laird 1999; Newstead & Evans 1995).
2. A syllogistic reasoning task in which logical validity conflicted with the believability of the conclusion (see Evans et al. 1983). An example item is: All mammals walk. Whales are mammals. Conclusion: Whales walk.
3. Statistical reasoning problems of the type studied by the Nisbett group (e.g., Fong et al. 1986) and inspired by the finding that human judgment is overly influenced by vivid but unrepresentative personal and case evidence and underinfluenced by more representative and diagnostic, but pallid, statistical evidence. The quintessential problem involves choosing between contradictory car purchase recommendations – one from a large-sample survey of car buyers and the other the heartfelt and emotional testimony of a single friend.
4. A covariation detection task modeled on the work of Wasserman et al. (1990). Subjects evaluated data derived from a 2 × 2 contingency matrix.
5. A hypothesis testing task modeled on Tschirgi (1980) in which the score on the task was the number of times subjects attempted to test a hypothesis in a manner that confounded variables.
6. A measure of outcome bias modeled on the work of 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.
7. A measure of if/only thinking bias (Epstein et al. 1992; Miller et al. 1990). If/only 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. 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.
8. An argument evaluation task (Stanovich & West 1997) that tapped reasoning skills of the type studied in the informal reasoning literature (Baron 1995; Klaczynski et al.

1997; Perkins et al. 1991). Important to see, it was designed so that to do well on it one had to adhere to a stricture not to implicate prior belief in the evaluation of the argument.

3. Computational limitations

Patterns of individual differences have implications that extend beyond testing the view that discrepancies between descriptive models and normative models arise entirely from performance errors. For example, patterns of individual differences also have implications for prescriptive models of rationality. Prescriptive models specify how reasoning should proceed given the limitations of the human cognitive apparatus and the situational constraints (e.g., time pressure) under which the decision maker operates (Baron 1985). Thus, normative models might not always be prescriptive for a given individual and situation. Judgments about the rationality of actions and beliefs must take into account the resource-limited nature of the human cognitive apparatus (Cherniak 1986; Goldman 1978; Harman 1995; Oaksford & Chater 1993; 1995; 1998; Stich 1990). More colloquially, Stich (1990) has argued that "it seems simply perverse to judge that subjects are doing a bad job of reasoning because they are not using a strategy that requires a brain the size of a blimp" (p. 27).

Following Dennett (1987) and the taxonomy of Anderson (1990; see also Marr 1982; Newell 1982), we distinguish the algorithmic/design level from the rational/intentional level of analysis in cognitive science (the first term in each pair is that preferred by Anderson, the second that preferred by Dennett). 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 of the system, beliefs relevant to those goals, and the choice of action that is rational given the system's goals and beliefs (Anderson 1990; Bratman et al. 1991; Dennett 1987; Newell 1982; 1990; Pollock 1995). However, even if all humans were optimally rational at the intentional level of analysis, there may still be computational limitations at the algorithmic level (e.g., Cherniak 1986; Goldman 1978; Oaksford & Chater 1993; 1995). We would, therefore, still expect individual differences in actual performance (despite equal rational-level competence) due to differences at the algorithmic level.

Using such a framework, we view the magnitude of the correlation between performance on a reasoning task and cognitive capacity as an empirical clue about the importance of algorithmic limitations in creating discrepancies between descriptive and normative models. A strong correlation suggests important algorithmic-level limitations that might make the normative response not prescriptive for those of lower cognitive capacity (Panglossian theorists drawn to this alternative explanation of normative/descriptive gaps were termed Apologists by Stanovich 1999.) 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).

In our studies, we have operationalized cognitive capacity in terms of well-known cognitive ability (intelligence) and academic aptitude tasks[2] but have most often used the total score on the Scholastic Assessment Test.[3,4] All are known to load highly on psychometric g (Carpenter et al. 1990; Carroll 1993; Matarazzo 1972), and such measures have been linked to neurophysiological and information-processing indicators of efficient cognitive computation (Caryl 1994; Deary 1995; Deary & Stough 1996; Detterman 1994; Fry & Hale 1996; Hunt 1987; Stankov & Dunn 1993; Vernon 1991; 1993). Furthermore, measures of general intelligence have been shown to be linked to virtually all of the candidate subprocesses of mentality that have been posited as determinants of cognitive capacity (Carroll 1993). For example, working memory is the quintessential component of cognitive capacity (in theories of computability, computational power often depends on memory for the results of intermediate computations). Consistent with this interpretation, Bara et al. (1995) have found that "as working memory improves – for whatever reason – it enables deductive reasoning to improve too" (p. 185). But it has been shown that, from a psychometric perspective, variation in working memory is almost entirely captured by measures of general intelligence (Kyllonen 1996; Kyllonen & Christal 1990).

Measures of general cognitive ability such as those used in our research are direct marker variables for Spearman's (1904; 1927) positive manifold – that performance on all reasoning tasks tends to be correlated. Below, we will illustrate how we use this positive manifold to illuminate reasons for the normative/descriptive gap.

Table 1 indicates the magnitude of the correlation between one such measure – Scholastic Assessment Test total scores – and the eight different reasoning tasks studied by Stanovich and West (1998c, Experiments 1 and 2) and mentioned in the previous section. In Experiment 1, syllogistic reasoning in the face of interfering content displayed the highest correlation (.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. Three of the four tasks from the previous experiment were carried over (all but the selection task) and displayed correlations similar in magnitude to those obtained in the first experiment. The correlations involving

Table 1. *Correlations between the reasoning tasks and Scholastic Assessment Test total scores in the Stanovich and West (1998c) studies*

Experiment 1	
Syllogisms	.470**
Selection task	.394**
Statistical reasoning	.347**
Argument evaluation task	.358**
Experiment 2	
Syllogisms	.410**
Statistical reasoning	.376**
Argument evaluation task	.371**
Covariation detection	.239**
Hypothesis testing bias	−.223**
Outcome bias	−.172**
If/only thinking	−.208**
Composite score	.547**

** = $p < .001$, all two-tailed

Ns = 178 to 184 in Experiment 1 and 527 to 529 in Experiment 2 the four new tasks introduced in Experiment 2 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. The correlations on the four new tasks were generally lower (.172 to .239) than the correlations involving the other tasks (.371 to .410). The scores on all of the tasks in Experiment 2 were standardized and summed to yield a composite score. The composite's correlation with Scholastic Assessment Test scores was .547. It thus appears that, to a moderate extent, discrepancies between actual performance and normative models can be accounted for by variation in computational limitations at the algorithmic level – at least with respect to the tasks investigated in these particular experiments.

However, there are some tasks in the heuristics and biases literature that lack any association at all with cognitive ability. The so-called false consensus effect in the opinion prediction paradigm (Krueger & Clement 1994; Krueger & Zeiger 1993) displays complete dissociation with cognitive ability (Stanovich 1999; Stanovich & West 1998c). Likewise, the overconfidence effect in the knowledge calibration paradigm (e.g., Lichtenstein et al. 1982) displays a negligible correlation with cognitive ability (Stanovich 1999; Stanovich & West 1998c).

Collectively, these results indicate that computational limitations seem far from absolute. That is, although computational limitations appear implicated to some extent in many of the tasks, the normative responses for all of them were computed by some university students who had modest cognitive abilities (e.g., below the mean in a university sample). Such results help to situate the relationship between prescriptive and normative models for the tasks in question because the boundaries of prescriptive recommendations for particular individuals might be explored by examining the distribution of the cognitive capacities of individuals who gave the normative response on a particular task. For most of these tasks, only a small number of the students with the very lowest cognitive ability in this sample would have prescriptive models that deviated substantially from the normative model for computational reasons. Such findings also might be taken to suggest that perhaps *other* factors might account for variation – a prediction that will be confirmed when work on styles of epistemic regulation is examined in section 7. Of course, the deviation between the normative and prescriptive model due to computational limitations will certainly be larger in unselected or nonuniversity populations. This point also serves to reinforce the caveat that the correlations observed in Table 1 were undoubtedly attenuated due to restriction of range in the sample. Nevertheless, if the normative/prescriptive gap is indeed modest, then there may well be true individual differences at the intentional level – that is, true individual differences in rational thought.

All of the camps in the dispute about human rationality recognize that positing computational limitations as an explanation for differences between normative and descriptive models is a legitimate strategy. Meliorists agree on the importance of assessing such limitations. Likewise, Panglossians will, when it is absolutely necessary, turn themselves into Apologists to rescue subjects from the charge of irrationality. Thus, they too acknowledge the importance of assessing computational limitations. In the next section, however, we examine an alternative explanation of the normative/descriptive gap that is much more controversial – the notion that inappropriate normative models have been applied to certain tasks in the heuristics and biases literature.

4. Applying the wrong normative model

The possibility of incorrect norm application arises because psychologists must appeal to the normative models of other disciplines (statistics, logic, etc.) in order to interpret the responses on various tasks, and these models must be applied to a particular problem or situation. 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:

> Many critics have insisted that in fact it is Kahneman & Tversky, not their subjects, who have failed to grasp the logic of the problem. (Margolis 1987, p. 158)
>
> If a "fallacy" is involved, it is probably more attributable to the researchers than to the subjects. (Messer & Griggs 1993, p. 195)
>
> When ordinary people reject the answers given by normative theories, they may do so out of ignorance and lack of expertise, or they may be signaling the fact that the normative theory is inadequate. (Lopes 1981, p. 344)
>
> In the examples of alleged base rate fallacy considered by Kahneman and Tversky, they, and not their experimental subjects, commit the fallacies. (Levi 1983, p. 502)
>
> What Wason and his successors judged to be the wrong response is in fact correct. (Wetherick 1993, p. 107)
>
> Perhaps the only people who suffer any illusion in relation to cognitive illusions are cognitive psychologists. (Ayton & Hardman 1997, p. 45)

These quotations reflect the numerous ongoing critiques of the heuristics and biases literature in which it is argued that the wrong normative standards have been applied to performance. For example, Lopes (1982) has argued that the literature on the inability of human subjects to generate random sequences (e.g., Wagenaar 1972) has adopted a narrow concept of randomness that does not acknowledge broader conceptions that are debated in the philosophy and mathematics literature. Birnbaum (1983) has 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 estimates from those assumed to be normatively correct under the less flexible Bayesian model that is usually applied. Gigerenzer (1991a; 1991b; 1993; Gigerenzer et al. 1991) has argued that the overconfidence effect in knowledge calibration experiments (Lichtenstein et al. 1982) and the conjunction effect in probability judgment (Tversky & Kahneman 1983) have been mistakenly classified as cognitive biases because of the application of an inappropriate normative model of probability assessment (i.e., requests for single-event subjective judgments, when under some conceptions of probability such judgments are not subject to the rules of a probability calculus). Dawes (1989; 1990) and Hoch (1987) have 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).

4.1. From the descriptive to the normative in reasoning and decision making

The cases just mentioned provide examples of how the existence of deviations between normative models and actual human reasoning has been called into question by casting doubt on the appropriateness of the normative models used to evaluate performance. Stein (1996, p. 239) terms this the "reject-the-norm" strategy. It is noteworthy that this strategy is frequently used by the Panglossian camp in the rationality debate, although this connection is not a necessary one. Specifically, Panglossians, exclusively used the reject-the-norm-application strategy to *eliminate* gaps between descriptive models of performance and normative models. When this type of critique is employed, the normative model that is suggested as a substitute for the one traditionally used in the heuristics and biases literature 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 ("Kahneman and Tversky . . . and not their experimental subjects, commit the fallacies"; Levi, 1983, p. 502), 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.

The fact that the use of the reject-the-norm-application strategy is entirely contingent on the existence or nonexistence of a normative/descriptive gap suggests that, at least for Panglossians, the strategy is empirically, not conceptually, triggered (normative applications are never rejected for purely conceptual reasons when they coincide with the modal human response). What this means is that in an important sense the applications of norms being endorsed by the Panglossian camp are conditioned (if not indexed entirely) 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.

Gigerenzer (1991b) is clear about his adherence to an empirically-driven reject-the-norm-application strategy:

> Since its origins in the mid-seventeenth century. . . . when there was a striking discrepancy between the judgment of reasonable men and what probability theory dictated – as with the famous St. Petersburg paradox – then the mathematicians went back to the blackboard and changed the equations (Daston 1980). Those good old days have gone. . . . If, in studies on social cognition, researchers find a discrepancy between human judgment and what probability theory seems to dictate, the blame is now put on the human mind, not the statistical model. (p. 109)

One way of framing the current debate between the Panglossians and Meliorists is to observe that the Panglossians wish for a return of the "good old days" where the normative was derived from the intuitions of the untutored layperson ("an appeal to people's intuitions is indispensable"; Cohen, 1981, p. 318); whereas the Meliorists (with their greater emphasis on the culturally constructed nature of norms) view the mode of operation during the "good old days" as a contingent fact of history – the product of a period when few aspects of epistemic and pragmatic rationality had been codified and preserved for general diffusion through education.

Thus, the Panglossian reject-the-norm-application view can in essence be seen as a conscious application of the naturalistic fallacy (deriving ought from is). For example, Cohen (1981), like Gigerenzer, feels that the normative is indexed to the descriptive in the sense that a competence model of actual behavior can simply be interpreted as the normative model. Stein (1996) notes that proponents of this position believe that the normative can simply be "read off" from a model of competence because "whatever human reasoning competence turns out to be, the principles embodied in it are the normative principles of reasoning" (p. 231). Although both endorse this linking of the normative to the descriptive, Gigerenzer (1991b) and Cohen (1981) do so for somewhat different reasons. For Cohen (1981), it follows from his endorsement of narrow reflective equilibrium as the sine qua non of normative justification. Gigerenzer's (1991b) endorsement is related to his position in the "cognitive ecologist" camp (to use Piattelli-Palmarini's 1994, p. 183 term) with its emphasis on the ability of evolutionary mechanisms to achieve an optimal Brunswikian tuning of the organism to the local environment (Brase et al. 1998; Cosmides & Tooby 1994; 1996; Oaksford & Chater 1994; 1998; Pinker 1997).

That Gigerenzer and Cohen concur here – even though they have somewhat different positions on normative justification – simply shows how widespread is the acceptance of the principle that descriptive facts about human behavior condition our notions about the appropriateness of the normative models used to evaluate behavior. In fact, stated in such broad form, this principle is not restricted to the Panglossian position. For example, in decision science, there is a long tradition of acknowledging descriptive influences when deciding which normative model to apply to a particular situation. Slovic (1995) refers to this "deep interplay between descriptive phenomena and normative principles" (p. 370). Larrick et al. (1993) have reminded us that "there is also a tradition of justifying, and amending, normative models in response to empirical considerations" (p. 332). March (1988) refers to this tradition when he discusses how actual human behavior has conditioned models of efficient problem solving in artificial intelligence and in the area of organizational decision making. The assumptions underlying the naturalistic project in epistemology (e.g., Kornblith 1985; 1993) have the same implication – that findings about how humans form and alter beliefs should have a bearing on which normative theories are correctly applied when evaluating the adequacy of belief acquisition. This position is in fact quite widespread:

> If people's (or animals') judgments do not match those predicted by a normative model, this may say more about the need for revising the theory to more closely describe subjects' cognitive processes than it says about the adequacy of those processes. (Alloy & Tabachnik 1984, p. 140)

> We must look to what people do in order to gather materials for epistemic reconstruction and self-improvement. (Kyburg 1991, p. 139)
>
> When ordinary people reject the answers given by normative theories, they may do so out of ignorance and lack of expertise, or they may be signaling the fact that the normative theory is inadequate. (Lopes 1981, p. 344)

Of course, in this discussion we have conjoined disparate views that are actually arrayed on a continuum. The reject-the-norm advocates represent the extreme form of this view – they simply want to read off the normative from the descriptive: "the argument under consideration here rejects the standard picture of rationality and takes the reasoning experiments as giving insight not just into human reasoning competence but also into the normative principles of reasoning" (Stein 1996, p. 233). In contrast, other theorists (e.g., March 1988) simply want to subtly fine-tune and adjust normative applications based on descriptive facts about reasoning performance.

One thing that all of the various camps in the rationality dispute have in common is that each conditions their beliefs about the appropriate norm to apply based on the *central tendency* of the responses to a problem. They all seem to see that single aspect of performance as the only descriptive fact that is relevant to conditioning their views about the appropriate normative model to apply. For example, advocates of the reject-the-norm-application strategy for dealing with normative/descriptive discrepancies view the mean, or modal, response as a direct pointer to the appropriate normative model. One goal of the present research program is to expand the scope of the descriptive information used to condition our views about appropriate applications of norms.

4.2. Putting descriptive facts to work: The understanding/acceptance assumption

How should we interpret situations where the majority of individuals respond in ways that depart from the normative model applied to the problem by reasoning experts? Thagard (1982) calls the two different interpretations the populist strategy and the elitist strategy: "The populist strategy, favored by Cohen (1981), is to emphasize the reflective equilibrium of the average person. . . . The elitist strategy, favored by Stich and Nisbett (1980), is to emphasize the reflective equilibrium of experts" (p. 39). Thus, Thagard (1982) identifies the populist strategy with the Panglossian position and the elitist strategy with the Meliorist position.

But there are few controversial tasks in the heuristics and biases literature where all untutored laypersons disagree with the experts. There are always some who agree. Thus, the issue is not the untutored average person versus experts (as suggested by Thagard's formulation), but experts plus some laypersons versus other untutored individuals. Might the cognitive characteristics of those departing from expert opinion have implications for which normative model we deem appropriate? Larrick et al. (1993) made just such an argument in their analysis of what justified the cost-benefit reasoning of microeconomics: "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.

Slovic and Tversky (1974) made essentially this argument years ago, although it was couched in very different terms in their paper and thus was hard to discern. Slovic and Tversky (1974) argued that descriptive facts about argument endorsement should condition the inductive inferences of experts regarding appropriate normative principles. In response to the argument that there is "no valid way to distinguish between outright rejection of the axiom and failure to understand it" (p. 372), Slovic and Tversky (1974) observed that "the deeper the understanding of the axiom, the greater the readiness to accept it" (pp. 372–73). Thus, a 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.

We might call Slovic and Tversky's argument the understanding/acceptance assumption – that more reflective and engaged reasoners are more likely to affirm the appropriate normative model for a particular situation. From their understanding/acceptance principle, it follows that if greater understanding resulted in more acceptance of the axiom, then the initial gap between the normative and descriptive would be attributed to factors that prevented problem understanding (e.g., lack of ability or reflectiveness on the part of the subject). Such a finding would increase confidence in the normative appropriateness of the axioms and/or in their application to a particular problem. In contrast, if better understanding failed to result in greater acceptance of the axiom, then its normative status for that particular problem might be considered to be undermined.

Using their understanding/acceptance principle, Slovic and Tversky (1974) examined the Allais (1953) problem and found little support for the applicability of the independence axiom of utility theory (the axiom stating that if the outcome in some state of the world is the same across options, then that state of the world should be ignored; Baron 1993; Savage 1954). When presented with arguments to explicate both the Allais (1953) and Savage (1954) positions, subjects found the Allais argument against independence at least as compelling and did not tend to change their task behavior in the normative direction (see MacCrimmon 1968; MacCrimmon & Larsson 1979 for more mixed results on the independence axiom using related paradigms). Although Slovic and Tversky (1974) failed to find support for this particular normative application, they presented a principle that may be of general usefulness in theoretical debates about why human performance deviates from normative models. The central idea behind Slovic and Tversky's (1974) development of the understanding/acceptance assumption is that increased understanding should drive performance in the direction of the truly normative principle for the particular situation – so that the direction that performance moves in response to increased understanding provides an empirical clue as to what is the proper normative model to be applied.

One might conceive of 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 employed the former strategy by providing subjects with explicated arguments supporting the Allais or Savage normative interpretation (see also Doherty et al. 1981; Stanovich & West 1999). Other methods of manipulating understanding have provided consistent evidence in favor of the normative principle of descriptive invariance (see Kahneman & Tversky 1984). For example, it has been found that being forced to take more time or to provide a rationale for selections increases adherence to descriptive invariance (Larrick et al. 1992; Miller & Fagley 1991; Sieck & Yates 1997; Takemura 1992; 1993; 1994). Moshman and Geil (1998) found that group discussion facilitated performance on Wason's selection task.

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; Kardash & Scholes 1996; Klaczynski et al. 1997; Smith & Levin 1996; Verplanken 1993). This particular application of the understanding/acceptance principle derives from the assumption that a normative/descriptive gap that is disproportionately created by subjects with a superficial understanding of the problem provides no warrant for amending the application of standard normative models.

4.3. Tacit acceptance of the understanding/acceptance principle as a mechanism for adjudicating disputes about the appropriate normative models to apply

It is important to point out that many theorists on all sides of the rationality debate have acknowledged the force of the understanding/acceptance argument (without always labeling the argument as such or citing Slovic & Tversky 1974). For example, Gigerenzer and Goldstein (1996) lament the fact that Apologists who emphasize Simon's (1956; 1957; 1983) concept of bounded rationality seemingly accept the normative models applied by the heuristics and biases theorists by their assumption that, if computational limitations were removed, individuals' responses would indeed be closer to the behavior those models prescribe.

Lopes and Oden (1991) also wish to deny this tacit assumption in the literature on computational limitations: "discrepancies between data and model are typically attributed to people's limited capacity to process information. . . . There is, however, no support for the view that people would choose in accord with normative prescriptions if they were provided with increased capacity" (pp. 208–209). In stressing the importance of the lack of evidence for the notion that people would "choose in accord with normative prescriptions if they were provided with increased capacity" (p. 209), Lopes and Oden (1991) acknowledge the force of the individual differences version of the understanding/acceptance principle – because examining variation in cognitive ability is just that: looking at what subjects who have "increased capacity" actually do with that increased capacity.

In fact, critics of the heuristics and biases literature have repeatedly drawn on an individual differences version of the understanding/acceptance principle to bolster their critiques. For example, Cohen (1982) critiques the older "bookbag and poker chip" literature on Bayesian conservatism (Phillips & Edwards 1966; Slovic et al. 1977) by noting that "if so-called 'conservatism' resulted from some inherent inadequacy in people's information-processing systems one might expect that, when individual differences in information-processing are measured on independently attested scales, some of them would correlate with degrees of 'conservatism.' In fact, no such correlation was found by Alker and Hermann (1971). And this is just what one would expect if 'conservatism' is not a defect, but a rather deeply rooted virtue of the system" (pp. 259–60). This is precisely how Alker and Hermann (1971) themselves argued in their paper:

> Phillips et al. (1966) have proposed that conservatism is the result of intellectual deficiencies. If this is the case, variables such as rationality, verbal intelligence, and integrative complexity should have related to deviation from optimality – more rational, intelligent, and complex individuals should have shown less conservatism. (p. 40)

Wetherick (1971; 1995) has been a critic of the standard interpretation of the four-card selection task (Wason 1966) for over 25 years. As a Panglossian theorist, he has been at pains to defend the modal response chosen by roughly 50% of the subjects (the P and Q cards). As did Cohen (1982) and Lopes and Oden (1991), Wetherick (1971) points to the lack of associations with individual differences to bolster his critique of the standard interpretation of the task: "in Wason's experimental situation subjects do not choose the not-Q card nor do they stand and give three cheers for the Queen, neither fact is interesting in the absence of a plausible theory predicting that they should. . . . If it could be shown that subjects who choose not-Q are more intelligent or obtain better degrees than those who do not this would make the problem worth investigation, but I have seen no evidence that this is the case" (Wetherick 1971, p. 213).

Funder (1987), like Cohen (1982) and Wetherick (1971), uses a finding about individual differences to argue that a particular attribution bias is not necessarily produced by a process operating suboptimally. Block and Funder (1986) analyzed the role effect observed by Ross et al. (1977): that people rated questioners more knowledgeable than contestants in a quiz game. Although the role effect is usually viewed as an attributional error – people allegedly failed to consider the individual's role when estimating the knowledge displayed – Block and Funder (1986) demonstrated that subjects most susceptible to this attributional "error" were more socially competent, more well adjusted, and more intelligent. Funder (1987) argued that "manifestation of this 'error,' far from being a symptom of social maladjustment, actually seems associated with a degree of competence" (p. 82) and that the so-called error is thus probably produced by a judgmental process that is generally efficacious. In short, the argument is that the signs of the correlations with the individual difference variables point in the direction of the response that is produced by processes that are ordinarily useful.

Thus, Funder (1987), Lopes and Oden (1991), Wetherick (1971), and Cohen (1982) all have recourse to patterns of individual differences (or the lack of such patterns) to pump our intuitions (Dennett 1980) in the direction of undermining the standard interpretations of the tasks under consideration. In other cases, however, examining individual differences may actually reinforce confidence in the appropriateness of the normative models applied to problems in the heuristics and biases literature.

4.4. The understanding/acceptance principle and Spearman's positive manifold

With these arguments in mind, it is thus interesting to note that the direction of all of the correlations displayed in Table 1 is consistent with the standard normative models used by psychologists working in the heuristics and biases tradition. The directionality of the systematic correlations with intelligence is embarrassing for those reject-the-norm-application theorists who argue that norms are being incorrectly applied if we interpret the correlations in terms of the understanding/acceptance principle (a principle which, as seen in sect. 4.3, is endorsed in various forms by a host of Panglossian critics of the heuristics and biases literature). Surely we would want to avoid the conclusion that individuals with more computational power are systematically computing the *non*normative response. Such an outcome would be an absolute first 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 one hundred years – had finally broken down. Obviously, parsimony dictates that positive manifold remains a fact of life for cognitive tasks and that the response originally thought to be normative actually is.

In fact, it is probably helpful to articulate the understanding/acceptance principle somewhat more formally in terms of 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 1991a; 1993; 1996a; 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.[5] Likewise, given that positive manifold is the norm among cognitive tasks, the negative correlation (or, to a lesser extent, the lack of a correlation) between a probabilistic reasoning task and more standard cognitive ability measures might be taken as a signal that the wrong normative model is being applied to the former task or that there are alternative models that are equally appropriate. The latter point is relevant because the pattern of results in our studies has *not* always mirrored the positive manifold displayed in Table 1. We have previously mentioned the false-consensus effect and overconfidence effect as such examples, and further instances are discussed in the next section.

4.5. Noncausal base rates

The statistical reasoning problems utilized in the experiments discussed so far (those derived from Fong et al. 1986) involved 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. Noncausal base-rate problems – those involving base rates with no obvious causal relationship to the criterion behavior – have had a much more controversial history in the research literature. They have been the subject of over a decade's worth of contentious dispute (Bar-Hillel 1990; Birnbaum 1983; Cohen 1979; 1982; 1986; Cosmides & Tooby 1996; Gigerenzer 1991b; 1993; 1996a; Gigerenzer & Hoffrage 1995; Kahneman & Tversky 1996; Koehler 1996; Kyburg 1983; Levi 1983; Macchi 1995) – important components of which have been articulated in this journal (e.g., Cohen 1981; 1983; Koehler 1996; Krantz 1981; Kyburg 1983; Levi 1983).

In several experiments, we have examined some of the noncausal base-rate problems that are notorious for provoking philosophical dispute. One was an AIDS testing problem modeled on Casscells et al. (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. In several analyses and replications (see Stanovich 1999; Stanovich & West 1998c) in which we have classified responses of less than 10% as Bayesian, responses of over 90% as indicating strong reliance on indicant information, and responses between 10% and 90% as intermediate, we have found that subjects giving the indicant response were higher in cognitive ability than those giving the Bayesian response.[6] Additionally, when tested on causal base-rate problems (e.g., Fong et al. 1986), the greatest base-rate usage was displayed by the group highly reliant on the *indicant* information in the AIDS problem. The subjects giving the Bayesian answer on the AIDS problem were least reliant on the aggregate information in the causal statistical reasoning problems.

A similar violation of the expectation of positive manifold was observed on the notorious cab problem (see Bar-Hillel 1980; Lyon & Slovic 1976; Tversky & Kahneman 1982) – also the subject of almost two decades-worth of dispute:

> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city in which the accident occurred. You are given the following facts: 85 percent of the cabs in the city are Green and 15 percent are Blue.

> A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each of the two colors 80 percent of the time. What is the probability that the cab involved in the accident was Blue?

Bayes's rule yields .41 as the posterior probability of the cab being blue. Thus, responses over 70% were classified as reliant on indicant information, responses between 30% and 70% as Bayesian, and responses less than 30% as reliant on base-rate information. Again, it was found that subjects giving the indicant response were higher in cognitive ability and need for cognition than those giving the Bayesian or base-rate response (Stanovich & West 1998c; 1999). Finally, both the cabs problem and the AIDS problem were subjected to the second of Slovic and Tversky's (1974) methods of operationalizing the understanding/acceptance principle – presenting the subjects with arguments explicating the traditional normative interpretation (Stanovich & West 1999). On neither problem was there a strong tendency for responses to move in the Bayesian direction subsequent to explication.

The results from both problems indicate that the noncausal base-rate problems display patterns of individual differences quite unlike those shown on the causal aggregate problems. On the latter, subjects giving the statistical response (choosing the aggregate rather than the case or indicant information) scored consistently higher on measures of cognitive ability. This pattern did not hold for the AIDS and cab problem where the significant differences were in the opposite direction – subjects strongly reliant on the *indicant* information scored higher on measures of cognitive ability and were *more* likely to give the Bayesian response on causal base-rate problems.

We examined the processing of noncausal base rates in another task with very different task requirements (see Stanovich 1999; Stanovich & West 1998d) – a selection task in which individuals were not forced to compute a Bayesian posterior, but instead simply had to indicate whether or not they thought the base rate was relevant to their decision. The task was taken from the work of Doherty and Mynatt (1990). Subjects were given the following instructions:

> Imagine you are a doctor. A patient comes to you with a red rash on his fingers. What information would you want in order to diagnose whether the patient has the disease Digirosa? Below are four pieces of information that may or may not be relevant to the diagnosis. Please indicate *all* of the pieces of information that are necessary to make the diagnosis, but *only* those pieces of information that are necessary to do so.

Subjects then chose from the alternatives listed in the order: percentage of people without Digirosa who have a red rash, percentage of people with Digirosa, percentage of people without Digirosa, and percentage of people with Digirosa who have a red rash. These alternatives represented the choices of P(D/~H), P(H), P(~H), and P(D/H), respectively.

The normatively correct choice of P(H), P(D/H), and P(D/~H) was made by 13.4% of our sample. The most popular choice (made by 35.5% of the sample) was the two components of the likelihood ratio, (P(D/H) and P(D/~H); 21.9% of the sample chose P(D/H) only; and 22.7% chose the base rate, P(H), and the numerator of the likelihood ratio, P(D/H) – ignoring the denominator of the likelihood ratio, P(D/~H). Collapsed across these combinations, almost all subjects (96.0%) viewed P(D/H) as relevant and very few (2.8%) viewed P(~H) as relevant. Overall, 54.3% of the subjects deemed that P(D/~H) was necessary information and 41.5% of the sample thought it was necessary to know the base rate, P(H).

We examined the cognitive characteristics of the subjects who thought the base rate was relevant and found that they did not display higher Scholastic Assessment Test scores than those who did not choose the base rate. The pattern of individual differences was quite different for the denominator of the likelihood ratio, P(D/~H) – a component that is normatively uncontroversial. Subjects seeing this information as relevant had significantly higher Scholastic Assessment Test scores.

Interesting to see, in light of these patterns of individual differences showing lack of positive manifold when the tasks are scored in terms of the standard Bayesian approach, noncausal base-rate problems like the AIDS and cab problems have been the focus of intense debate in the literature (Cohen 1979; 1981; 1982; 1986; Koehler 1996; Kyburg 1983; Levi 1983). Several authors have argued that a rote application of the Bayesian formula to these problems is unwarranted because noncausal base rates of the AIDS-problem type lack relevance and reference-class specificity. Finally, our results might also suggest that the Bayesian subjects on the AIDS problem might not actually be arriving at their response through anything resembling Bayesian processing (whether or not they were operating in a frequentist mode; Gigerenzer & Hoffrage 1995), because on causal aggregate statistical reasoning problems these subjects were less likely to rely on the aggregate information.

5. Alternative task construals

Theorists who resist interpreting the gap between normative and descriptive models as indicating human irrationality have one more strategy available in addition to those previously described. In the context of empirical cognitive psychology, it is a commonplace argument, but it is one that continues to create enormous controversy and to bedevil efforts to compare human performance to normative standards. It 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, e.g., Adler 1984; 1991; Broome 1990; Henle 1962; Hilton 1995; Levinson 1995; Margolis 1987; Schick 1987; 1997; Schwarz 1996).

Such an argument is somewhat different from any of the critiques examined thus far. It is not the equivalent of positing that a performance error has been made, because performance errors (attention lapses, etc.) – being transitory and random – would not be expected to recur in exactly the same way in a readministration of the same task. Whereas, if the subject has truly misunderstood the task, they would be expected to do so again on an identical readministration of the task.

Correspondingly, this criticism is different from the argument that the task exceeds the computational capacity of the subject. The latter explanation locates the cause of the suboptimal performance within the subject. 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 subjects to frame the problem in a manner different from that intended.[7]

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.

It seems that in order to comprehensively evaluate the rationality of human cognition it will be necessary to evaluate the appropriateness of various task construals. This is because – contrary to thin theories of means/ends rationality that avoid evaluating the subject's task construal (Elster 1983; Nathanson 1994) – it will be argued here that if we are going to have any normative standards at all, then we must also have standards for what are appropriate and inappropriate task construals. In the remainder of this section, we will sketch the arguments of philosophers and decision scientists who have made just this point. Then it will be argued that:

1. In order to tackle the difficult problem of evaluating task construals, criteria of wide reflective equilibrium come into play;
2. It will be necessary to use all descriptive information about human performance that could potentially affect expert wide reflective equilibrium;
3. Included in the relevant descriptive facts are individual differences in task construal and their patterns of covariance. This argument will again make use of the understanding/acceptance principle of Slovic and Tversky (1974) discussed in section 4.2.

5.1. The necessity of principles of rational construal

It is now widely recognized that the evaluation of the normative appropriateness of a response to a particular task is always relative to a particular interpretation of the task. For example, Schick (1987) argues that "how rationality directs us to choose depends on which understandings are ours . . . [and that] the understandings people have bear on the question of what would be rational for them" (pp. 53 and 58). Likewise, Tversky (1975) argued that "the question of whether utility theory is compatible with the data or not, therefore, depends critically on the interpretation of the consequences" (p. 171).

However, others have pointed to the danger inherent in too permissively explaining away nonnormative responses by positing different construals of the problem. Normative theories will be drained of all of their evaluative force if we adopt an attitude that is too charitable toward alternative construals. Broome (1990) illustrates the problem by discussing the preference reversal phenomenon (Lichtenstein & Slovic 1971; Slovic 1995). In a choice between two gambles, A and B, a person chooses A over B. However, when pricing the gambles, the person puts a higher price on B. This violation of procedural invariance leads to what appears to be intransitivity. Presumably there is an amount of money, M, that would be preferred to A but given a choice of M and B the person would choose B. Thus, we appear to have $B > M$, $M > A$, $A > B$. Broome (1990) points out that when choosing A over B the subject is choosing A and is simultaneously rejecting B. Evaluating A in the M versus A comparison is not the same. Here, when choosing A, the subject is not rejecting B. The A alternative here might be considered to be a different prospect (call it A'), and if it is so considered there is no intransitivity ($B > M$, $M > A'$, $A > B$). Broome (1990) argues that whenever the basic axioms such as transitivity, independence, or descriptive or procedural invariance are breached, the same inoculating strategy could be invoked – that of individuating outcomes so finely that the violation disappears.

Broome's (1990) point is that the thinner the categories we use to individuate outcomes, the harder it will be to attribute irrationality to a set of preferences if we evaluate rationality only in instrumental terms. He argues that we need, in addition to the formal principles of rationality, principles that deal with *content* so as to enable us to evaluate the reasonableness of a particular individuation of outcomes. Broome (1990) acknowledges that "this procedure puts principles of rationality to work at a very early stage of decision theory. They are needed in fixing the set of alternative prospects that preferences can then be defined upon. The principles in question might be called "rational principles of indifference" (p. 140). Broome (1990) admits that

> many people think there can be no principles of rationality apart from the formal ones. This goes along with the common view that rationality can only be instrumental . . . [however] if you acknowledge only formal principles of rationality, and deny that there are any principles of indifference, you will find yourself without any principles of rationality at all. (pp. 140–41)

Broome cites Tversky (1975) as concurring in this view:

> I believe that an adequate analysis of rational choice cannot accept the evaluation of the consequences as given, and examine only the consistency of preferences. There is probably as much irrationality in our feelings, as expressed in the way we evaluate consequences, as there is in our choice of actions. An adequate normative analysis must deal with problems such as the legitimacy of regret in Allais' problem. . . . I do not see how the normative appeal of the axioms could be discussed without a reference to a specific interpretation. (Tversky 1975, p. 172)

Others agree with the Broome/Tversky analysis (see Baron 1993; 1994; Frisch 1994; Schick 1997). Although there is some support for Broome's generic argument, the contentious disputes about rational principles of indifference and rational construals of the tasks in the heuristics and biases literature (Adler 1984; 1991; Berkeley & Humphreys 1982; Cohen 1981; 1986; Gigerenzer 1993; 1996a; Hilton 1995; Jepson et al. 1983; Kahneman & Tversky 1983; 1996; Lopes 1991; Nisbett 1981; Schwarz 1996) highlight the difficulties to be faced when attempting to evaluate specific problem construals. For example, Margolis (1987) agrees with Henle (1962) that the subjects' nonnormative responses will almost always be logical responses to some other problem representation. But unlike Henle (1962), Margolis (1987) argues that many of these alternative task construals are so bizarre – so far from what the very words in the instructions said – that they represent serious cognitive errors that deserve attention:

> But in contrast to Henle and Cohen, the detailed conclusions I draw strengthen rather than invalidate the basic claim of the ex-

> perimenters. For although subjects can be – in fact, I try to show, ordinarily are – giving reasonable responses to a different question, the different question can be wildly irrelevant to anything that plausibly could be construed as the meaning of the question asked. The locus of the illusion is shifted, but the force of the illusion is confirmed not invalidated or explained away. (p. 141)

5.2. Evaluating principles of rational construal: The understanding/acceptance assumption revisited

Given current arguments that principles of rational construal are necessary for a full normative theory of human rationality (Broome 1990; Einhorn & Hogarth 1981; Jungermann 1986; Schick 1987; 1997; Shweder 1987; Tversky 1975), how are such principles to be derived? When searching for principles of rational task construal the same mechanisms of justification used to assess principles of instrumental rationality will be available. Perhaps in some cases – instances where the problem structure maps the world in an unusually close and canonical way – problem construals could be directly evaluated by how well they serve the decision makers in achieving their goals (Baron 1993; 1994). In such cases, it might be possible to prove the superiority or inferiority of certain construals by appeals to Dutch Book or money pump arguments (de Finetti 1970/1990; Maher 1993; Osherson 1995; Resnik 1987; Skyrms 1986).

Also available will be the expert wide reflective equilibrium view discussed by Stich and Nisbett (1980) (see Stanovich 1999; Stein 1996). In contrast, Baron (1993; 1994) and Thagard (1982) argue that rather than any sort of reflective equilibrium, what is needed here are "arguments that an inferential system is optimal with respect to the criteria discussed" (Thagard 1982, p. 40). But in the area of task construal, finding optimization of criteria may be unlikely – there will be few money pumps or Dutch Books to point the way. If in the area of task construal there will be few money pumps or Dutch Books to prove that a particular task interpretation has disastrous consequences, then the field will be again thrust back upon the debate that Thagard (1982) calls "the argument between the populists and the elitists." But as argued before, this is really a misnomer. There are few controversial tasks in the heuristics and biases literature where *all* untutored laypersons interpret tasks differently from the experts who designed them. The issue is not the untutored average person versus experts, but experts plus some laypersons versus other untutored individuals. The cognitive characteristics of those departing from the expert construal might – for reasons parallel to those argued in section 4 – have implications for how we evaluate particular task interpretations. It is argued here that Slovic and Tversky's (1974) assumption ("the deeper the understanding of the axiom, the greater the readiness to accept it" pp. 372–73) can again be used as a tool to condition the expert reflective equilibrium regarding principles of rational task construal.

Framing effects are ideal vehicles for demonstrating how the understanding/acceptance principle might be utilized. First, it has already been shown that there are consistent individual differences across a variety of framing problems (Frisch 1993). Second, framing problems have engendered much dispute regarding issues of appropriate task construal. The Disease Problem of Tversky and Kahneman (1981) has been the subject of much contention:

Problem 1. Imagine that the United States is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows: If Program A is adopted, 200 people will be saved. If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved. Which of the two programs would you favor, Program A or Program B?

Problem 2. Imagine that the United States is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows: If Program C is adopted, 400 people will die. If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. Which of the two programs would you favor, Program C or Program D?

Many subjects select alternatives A and D in these two problems despite the fact that the two problems are redescriptions of each other and that Program A maps to Program C rather than D. This response pattern violates the assumption of descriptive invariance of utility theory. However, Berkeley and Humphreys (1982) argue that Programs A and C might not be descriptively invariant in subjects' interpretations. They argue that the wording of the outcome of Program A ("will be saved") combined with the fact that its outcome is seemingly not described in the exhaustive way as the consequences for Program B suggests the possibility of human agency in the future which might enable the saving of *more* lives (see Kuhberger 1995). The wording of the outcome of Program C ("will die") does not suggest the possibility of future human agency working to possibly save more lives (indeed, the possibility of *losing* a few more might be inferred by some people). Under such a construal of the problem, it is no longer nonnormative to choose Programs A and D. Likewise, Macdonald (1986) argues that, regarding the "200 people will be saved" phrasing, "it is unnatural to predict an exact number of cases" (p. 24) and that "ordinary language reads 'or more' into the interpretation of the statement" (p. 24; see also Jou et al. 1996).

However, consistent with the finding that being forced to provide a rationale or take more time reduces framing effects (e.g., Larrick et al. 1992; Sieck & Yates 1997; Takemura 1994) and that people higher in need for cognition displayed reduced framing effects (Smith & Levin 1996), in our within-subjects study of framing effects on the Disease Problem (Stanovich & West 1998b), we found that subjects giving a consistent response to both descriptions of the problem – who were actually the majority in our within-subjects experiment – were significantly higher in cognitive ability than those subjects displaying a framing effect. Thus, the results of studies investigating the effects of giving a rationale, taking more time, associations with cognitive engagement, and associations with cognitive ability are all consistent in suggesting that the response dictated by the construal of the problem originally favored by Tversky and Kahneman (1981) should be considered the correct response because it is endorsed even by untutored subjects as long as they are cognitively engaged with the problem, had enough time to process the information, and had the cognitive ability to fully process the information.[8]

Perhaps no finding in the heuristics and biases literature has been the subject of as much criticism as Tversky and Kahneman's (1983) claim to have demonstrated a conjunction fallacy in probabilistic reasoning. Most of the criticisms have focused on the issue of differential task construal, and several critics have argued that there are alternative construals of the tasks that are, if anything, more rational than that which Tversky and Kahneman (1983) regard as normative for examples such as the well-known Linda Problem:

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 antinuclear 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 c or f, yet 85% of the subjects in Tversky and Kahneman's (1983) study rated alternative h as more probable than f. What concerns us here is the argument that there are subtle linguistic and pragmatic features of the problem which lead subjects to evaluate alternatives different from those listed. For example, Hilton (1995) argues 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 h as more probable than f no longer represents a conjunction fallacy.

Similarly, Morier and Borgida (1984) point out that the presence of the unusual conjunction "Linda is a bank teller and is active in the feminist movement" itself might prompt an interpretation of "Linda is a bank teller" as "Linda is a bank teller and is not active in the feminist movement." Actually, Tversky and Kahneman (1983) themselves had concerns about such an interpretation of the "Linda is a bank teller" alternative and ran a condition in which this alternative was rephrased as "Linda is a bank teller whether or not she is active in the feminist movement." They found that the conjunction fallacy was reduced from 85% of their sample to 57% when this alternative was used. Several other investigators have suggested that pragmatic inferences lead to seeming violations of the logic of probability theory in the Linda Problem[9] (see Adler 1991; Dulany & Hilton 1991; Levinson 1995; Macdonald & Gilhooly 1990; Politzer & Noveck 1991; Slugoski & Wilson 1998). These criticisms all share the implication that actually committing the conjunction fallacy is a rational response to an alternative construal of the different statements about Linda.

Assuming that those committing the so-called conjunction fallacy are making the pragmatic interpretation and that those avoiding the fallacy are making the interpretation that the investigators intended, we examined whether the subjects making the pragmatic interpretation were subjects who were disproportionately the subjects 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) – it might be expected that these individuals would be the subjects of higher cognitive ability.

In our study (Stanovich & West 1998b), we examined the performance of 150 subjects on the Linda Problem presented above. Consistent with the results of previous experiments on this problem (Tversky & Kahneman 1983), 80.7% of our sample displayed the conjunction effect – they rated the feminist bank teller alternative as more probable than the bank teller alternative. The mean SAT score of the 121 subjects who committed the conjunction fallacy was 82 points lower than the mean score of the 29 who avoided the fallacy. This difference was highly significant and it translated into an effect size of .746, which Rosenthal and Rosnow (1991, p. 446) classify as "large."

Tversky and Kahneman (1983) and Reeves and Lockhart (1993) have demonstrated that the incidence of the conjunction fallacy can be decreased if the problem describes the event categories in some finite population or if the problem is presented in a frequentist manner (see also Fiedler 1988; Gigerenzer 1991b; 1993). We have replicated this well-known finding, but we have also found that frequentist representations of these problems markedly reduce – if not eliminate – cognitive ability differences (Stanovich & West 1998b).

Another problem that has spawned many arguments about alternative construals is Wason's (1966) selection task. Performance on abstract versions of the selection task is extremely low (see Evans et al. 1993). Typically, less than 10% of subjects make the correct selections of the A card (P) and 7 card (not-Q). The most common incorrect choices made by subjects 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 & Lynch 1973), but some investigators have championed an explanation based on an alternative task construal. For example, Oaksford and Chater (1994; 1996; see also Nickerson 1996) argue that rather than interpreting the task as one of deductive reasoning (as the experimenter intends), many subjects interpret it as an inductive problem of probabilistic hypothesis testing. They show that the P and Q response is expected under a formal Bayesian analysis which assumes such an interpretation in addition to optimal data selection.

We have examined individual differences in responding on a variety of abstract and deontic selection task problems (Stanovich & West 1998a; 1998c). Typical results are displayed in Table 2. The table presents the mean Scholastic Assessment Test scores of subjects responding correctly (as traditionally interpreted – with the responses P and not-Q) on various versions of selection task problems. One was a commonly used nondeontic problem with content, the so-called Destination Problem (e.g., Manktelow & Evans 1979). Replicating previous research, few subjects responded correctly on this problem. However, those that did had significantly higher Scholastic Assessment Test scores

Table 2. *Mean Scholastic Assessment Test (SAT) total scores of subjects who gave the correct and incorrect responses to three different selection task problems (numbers in parentheses are the number of subjects)*

	Incorrect	P & not-Q (Correct)	t value	Effect size[a]
Nondeontic problem:				
Destination Problem	1,187 (197)	1,270 (17)	3.21***	.815
Deontic Problems:				
Drinking-Age Problem	1,170 (72)	1,206 (143)	2.39**	.347
Sears Problem	1,189 (87)	1,198 (127)	0.63	.088
	P & Q	**P & not-Q**	**t value**	**Effect size[a]**
Nondeontic Problem:				
Destination Problem	1,195 (97)	1,270 (17)	3.06***	.812

Note: df = 212 for the Destination and Sears Problems and 213 for the Drinking-Age Problem; df = 112 for the P&Q comparison on the Destination Problem

* = $p < .05$, ** = $p < .025$, *** = $p < .01$, all two-tailed

[a] Cohen's d

than those that did not and the difference was quite large in magnitude (effect size of .815). Also presented in the table are two well-known problems (Dominowski 1995; Griggs 1983; Griggs & Cox 1982; 1983; Newstead & Evans 1995) with deontic rules (reasoning about rules used to guide human behavior – about what "ought to" or "must" be done, see Manktelow & Over 1991) – the Drinking-Age Problem (If a person is drinking beer then the person must be over 21 years of age.) and the Sears Problem (Any sale over $30 must be approved by the section manager, Mr. Jones.). Both are known to facilitate performance and this effect is clearly replicated in the data presented in Table 2. However, it is also clear that the differences in cognitive ability are much less in these two problems. The effect size is reduced from .815 to .347 in the case of the Drinking-Age Problem and it fails to even reach statistical significance in the case of the Sears Problem (effect size of .088). The bottom half of the table indicates that exactly the same pattern was apparent when the P and not-Q responders were compared only with the P and Q responders on the Destination Problem – the latter being the response that is most consistent with an inductive construal of the problem (see Nickerson 1996; Oaksford & Chater 1994; 1996).

Thus, on the selection task, it appears that cognitive ability differences are strong in cases where there is a dispute about the proper construal of the task (in nondeontic tasks). In cases where there is little controversy about alternative construals – the deontic rules of the Drinking-Age and Sears problems – cognitive ability differences are markedly attenuated. This pattern – cognitive ability differences large on problems where there is contentious dispute regarding the appropriate construal and cognitive ability differences small when there is no dispute about task construal – is mirrored in our results on the conjunction effect and framing effect (Stanovich & West 1998b).

6. Dual process theories and alternative task construals

The sampling of results just presented (for other examples, see Stanovich 1999) has demonstrated that the responses associated with alternative construals of a well-known framing problem (the Disease Problem), of the Linda Problem, and of the nondeontic selection task were consistently associated with lower cognitive ability. How might we interpret this consistent pattern displayed on three tasks from the heuristics and biases literature where alternative task construals have been championed?

One possible interpretation of this pattern is in terms of two-process theories of reasoning (Epstein 1994; Evans 1984; 1996; Evans & Over 1996; Sloman 1996). A summary of the generic properties distinguished by several two-process views is presented in Table 3. Although the details and technical properties of these dual-process theories do not always match exactly, nevertheless there are clear family resemblances (for discussions, see Evans & Over 1996; Gigerenzer & Regier 1996; Sloman 1996). In order to emphasize the prototypical view that is adopted here, the two systems have simply been generically labeled System 1 and System 2.

The key differences in the properties of the two systems are listed next. System 1 is characterized as automatic, largely unconscious, and relatively undemanding of computational capacity. Thus, it conjoins properties of automaticity and heuristic processing as these constructs have been variously discussed in the literature. These properties characterize what Levinson (1995) has termed interactional intelligence – a system composed of the mechanisms that support a Gricean theory of communication that relies on intention-attribution. This system has as its goal the ability to model other minds in order to read intention and to make rapid interactional moves based on those modeled intentions. System 2 conjoins the various characteristics that have been viewed as typifying controlled processing. System 2 encompasses the processes of analytic intelligence that have traditionally been studied by information processing theorists trying to uncover the computational components underlying intelligence.

For the purposes of the present discussion, the most important difference between the two systems is that they tend to lead to different types of task construals. Construals triggered by System 1 are highly contextualized, per-

Table 3. *The terms for the two systems used by a variety of theorists and the properties of dual-process theories of reasoning*

	System 1	System 2
Dual-Process Theories:		
Sloman (1996)	associative system	rule-based system
Evans (1984;1989)	heuristic processing	analytic processing
Evans & Over (1996)	tacit thought processes	explicit thought processes
Reber (1993)	implicit cognition	explicit learning
Levinson (1995)	interactional intelligence	analytic intelligence
Epstein (1994)	experiential system	rational system
Pollock (1991)	quick and inflexible modules	intellection
Hammond (1996)	intuitive cognition	analytical cognition
Klein (1998)	recognition-primed decisions	rational choice strategy
Johnson-Laird (1983)	implicit inferences	explicit inferences
Shiffrin & Schneider (1977)	automatic processing	controlled processing
Posner & Snyder (1975)	automatic activation	conscious processing system
Properties:	associative holistic automatic relatively undemanding of cognitive capacity relatively fast acquisition by biology, exposure, and personal experience	rule-based analytic controlled demanding of cognitive capacity relatively slow acquisition by cultural and formal tuition
Task Construal	highly contextualized personalized conversational and socialized	decontextualized depersonalized asocial
Type of Intelligence Indexed:	interactional (conversational implicature)	analytic (psychometric IQ)

sonalized, and socialized. They are driven by considerations of relevance and are aimed at inferring intentionality by the use of conversational implicature even in situations that are devoid of conversational features (see Margolis 1987). The primacy of these mechanisms leads to what has been termed the fundamental computational bias in human cognition (Stanovich 1999) – the tendency toward automatic contextualization of problems. In contrast, System 2's more controlled processes serve to decontextualize and depersonalize problems. This system is more adept at representing in terms of rules and underlying principles. It can deal with problems without social content and is not dominated by the goal of attributing intentionality nor by the search for conversational relevance.

Using the distinction between System 1 and System 2 processing, it is conjectured here that in order to observe large cognitive ability differences in a problem situation, the two systems must strongly cue *different* responses.[10] It is not enough simply that both systems are engaged. If both cue the same response (as in deontic selection task problems), then this could have the effect of severely diluting any differences in cognitive ability. One reason that this outcome is predicted is that it is assumed that individual differences in System 1 processes (interactional intelligence) bear little relation to individual differences in System 2 processes (analytic intelligence). This is a conjecture for which there is a modest amount of evidence. Reber (1993) has shown preconscious processes to have low variability and to show little relation to analytic intelligence (see Jones & Day 1997; McGeorge et al. 1997; Reber et al. 1991).

In contrast, if the two systems cue opposite responses, rule-based System 2 will tend to differentially cue those of high analytic intelligence and this tendency will not be diluted by System 1 (the associative system) nondifferentially drawing subjects to the same response. For example, the Linda Problem maximizes the tendency for the two systems to prime different responses and this problem produced a large difference in cognitive ability. Similarly, in nondeontic selection tasks there is ample opportunity for the two systems to cue different responses. A deductive interpretation conjoined with an exhaustive search for falsifying instances yields the response P and not-Q. This interpretation and processing style is likely associated with the rule-based System 2 – individual differences which underlie the psychometric concept of analytic intelligence. In contrast, within the heuristic-analytic framework of Evans (1984; 1989; 1996), the matching response of P and Q reflects the heuristic processing of System 1 (in Evans' theory, a linguistically cued relevance response).

In deontic problems, both deontic and rule-based logics are cuing construals of the problem that dictate the same response (P and not-Q). Whatever is one's theory of responding in deontic tasks – preconscious relevance judgments, pragmatic schemas, or Darwinian algorithms (e.g., Cheng & Holyoak 1989; Cosmides 1989; Cummins 1996; Evans 1996) – the mechanisms triggering the correct response resemble heuristic or modular structures that fall within the domain of System 1. These structures are unlikely to be strongly associated with analytic intelligence (Cummins 1996; Levinson 1995; McGeorge et al. 1997; Reber 1993; Reber et al. 1991), and hence they operate to draw subjects of *both* high and low analytic intelligence to the same response dictated by the rule-based system – thus serving to dilute cognitive ability differences between correct and incorrect responders (see Stanovich & West 1998a for a data simulation).

6.1. Alternative construals: Evolutionary optimization versus normative rationality

The sampling of experimental results reviewed here (see Stanovich 1999 for further examples) has demonstrated that the response dictated by the construal of the inventors of the Linda Problem (Tversky & Kahneman 1983), Disease Problem (Tversky & Kahneman 1981), and selection task (Wason 1966) is the response favored by subjects of high analytic intelligence. The alternative responses dictated by the construals favored by the critics of the heuristics and biases literature were the choices of the subjects of lower analytic intelligence. In this section we will explore the possibility that these alternative construals may have been triggered by heuristics that make evolutionary sense, but that subjects higher in a more flexible type of analytic intelligence (and those more cognitively engaged, see Smith & Levin 1996) are more prone to follow normative rules that maximize personal utility. In a very restricted sense, such a pattern might be said to have relevance for the concept of rational task construal.

The argument depends on the distinction between evolutionary adaptation and instrumental rationality (utility maximization given goals and beliefs). The key point is that for the latter (variously termed practical, pragmatic, or means/ends rationality), maximization is at the level of the individual person. Adaptive optimization in the former case is at the level of the genes. In Dawkins's (1976; 1982) terms, evolutionary adaptation concerns optimization processes relevant to the so-called replicators (the genes), whereas instrumental rationality concerns utility maximization for the so-called vehicle (or interactor, to use Hull's 1982 term), which houses the genes. Anderson (1990; 1991) emphasizes this distinction in his treatment of adaptationist models in psychology. In his advocacy of such models, Anderson (1990; 1991) eschews Dennett's (1987) assumption of perfect rationality in the instrumental sense (hereafter termed normative rationality) for the somewhat different assumption of evolutionary optimization (i.e., evolution as a local fitness maximizer). Anderson (1990) accepts Stich's (see also Cooper 1989; Skyrms 1996) argument that evolutionary adaptation (hereafter termed evolutionary rationality)[11] does not guarantee perfect human rationality in the normative sense:

> Rationality in the adaptive sense, which is used here, is not rationality in the normative sense that is used in studies of decision making and social judgment. . . . It is possible that humans are rational in the adaptive sense in the domains of cognition studied here but not in decision making and social judgment. (1990, p. 31)

Thus, Anderson (1991) acknowledges that there may be arguments for "optimizing money, the happiness of oneself and others, or any other goal. It is just that these goals do not produce optimization of the species" (pp. 510–11). As a result, a descriptive model of processing that is adaptively optimal could well deviate substantially from a normative model. This is because Anderson's (1990; 1991) adaptation assumption is that cognition is optimally adapted in an *evolutionary* sense – and this is not the same as positing that human cognitive activity will result in normatively appropriate responses.

Such a view can encompass both the impressive record of descriptive accuracy enjoyed by a variety of adaptationist models (Anderson 1990; 1991; Oaksford & Chater 1994; 1996; 1998) as well as the fact that cognitive ability sometimes dissociates from the response deemed optimal on an adaptationist analysis (Stanovich & West 1998a). As discussed above, Oaksford and Chater (1994) have had considerable success in modeling the nondeontic selection task as an inductive problem in which optimal data selection is assumed (see also Oaksford et al. 1997). Their model predicts the modal response of P and Q and the corresponding dearth of P and not-Q choosers. Similarly, Anderson (1990, p. 157–60) models the 2 × 2 contingency assessment experiment using a model of optimally adapted information processing and shows how it can predict the much-replicated finding that the D cell (cause absent and effect absent) is vastly underweighted (see also Friedrich 1993; Klayman & Ha 1987). Finally, a host of investigators (Adler 1984; 1991; Dulany & Hilton 1991; Hilton 1995; Levinson 1995) have stressed how a model of rational conversational implicature predicts that violating the conjunction rule in the Linda Problem reflects the adaptive properties of interactional intelligence.

Yet in all three of these cases – despite the fact that the adaptationist models predict the modal response quite well – individual differences analyses demonstrate associations that also must be accounted for. Correct responders on the nondeontic selection task (P and not-Q choosers – not those choosing P and Q) are higher in cognitive ability. In the 2 × 2 covariation detection experiment, it is those subjects weighting cell D more equally (not those underweighting the cell in the way that the adaptationist model dictates) who are higher in cognitive ability and who tend to respond normatively on other tasks (Stanovich & West 1998d). Finally, despite conversational implicatures indicating the opposite, individuals of higher cognitive ability disproportionately tend to adhere to the conjunction rule. These patterns make sense if it is assumed that the two systems of processing are optimized for different situations and different goals and that these data patterns reflect the greater probability that the analytic intelligence of System 2 will override the interactional intelligence of System 1 in individuals of higher cognitive ability.

In summary, the biases introduced by System 1 heuristic processing may well be universal – because the computational biases inherent in this system are ubiquitous and shared by all humans. However, it does not necessarily follow that errors on tasks from the heuristics and biases literature will be universal (we have known for some time that they are not). This is because, for some individuals, System 2 processes operating in parallel (see Evans & Over 1996) will have the requisite computational power (or a low enough threshold) to override the response primed by System 1.

It is hypothesized that the features of System 1 are designed to very closely track increases in the reproduction probability of genes. System 2, while also clearly an evolutionary product, is also primarily a control system focused on the interests of the whole person. It is the primary maximizer of an individual's personal utility.[12] Maximizing the latter will occasionally result in sacrificing genetic fitness (Barkow 1989; Cooper 1989; Skyrms 1996). Because System 2 is more attuned to normative rationality than is System 1, System 2 will seek to fulfill the individual's goals in the minority of cases where those goals conflict with the responses triggered by System 1.

It is proposed that just such conflicts are occurring in three of the tasks discussed previously (the Disease Problem, the Linda Problem, and the selection task). This conjecture is supported by the fact that evolutionary rationality has been conjoined with Gricean principles of conversational implicature by several theorists (Gigerenzer 1996b; Hilton 1995; Levinson 1995) who emphasize the principle of "conversationally rational interpretation" (Hilton 1995, p. 265). According to this view, the pragmatic heuristics are not simply inferior substitutes for computationally costly logical mechanisms that would work better. Instead, the heuristics are optimally designed to solve an evolutionary problem in another domain – attributing intentions to conspecifics and coordinating mutual intersubjectivity so as to optimally negotiate cooperative behavior (Cummins 1996; Levinson 1995; Skyrms 1996).

It must be stressed though that in the vast majority of mundane situations, the evolutionary rationality embodied in System 1 processes will *also* serve the goals of normative rationality. Our automatic, System 1 processes for accurately navigating around objects in the natural world were adaptive in an evolutionary sense, and they likewise serve our personal goals as we carry out our lives in the modern world (i.e., navigational abilities are an evolutionary adaptation that serve the instrumental goals of the vehicle as well).

One way to view the difference between what we have termed here evolutionary and normative rationality is to note that they are not really two different *types* of rationality (see Oaksford & Chater 1998, pp. 291–97) but are instead terms for characterizing optimization procedures operating at the subpersonal and personal levels, respectively. That there are two optimization procedures in operation here that could come into conflict is a consequence of the insight that the genes – as subpersonal replicators – can increase their fecundity and longevity in ways that do not necessarily serve the instrumental goals of the vehicles built by the genome (Cooper 1989; Skyrms 1996).

Skyrms (1996) devotes an entire book on evolutionary game theory to showing that the idea that "natural selection will weed out irrationality" (p. x) is false because optimization at the subpersonal replicator level is not coextensive with the optimization of the instrumental goals of the vehicle (i.e., normative rationality). Gigerenzer (1996b) provides an example by pointing out that neither rats nor humans maximize utility in probabilistic contingency experiments. Instead of responding by choosing the most probable alternative on every trial, subjects alternate in a manner that matches the probabilities of the stimulus alternatives. This behavior violates normative strictures on utility maximization, but Gigerenzer (1996b) demonstrates how probability matching could actually be an evolutionarily stable strategy (see Cooper 1989 and Skyrms 1996 for many such examples).

Such examples led Skyrms (1996) to note that "when I contrast the results of the evolutionary account with those of rational decision theory, I am not criticizing the normative force of the latter. I am just emphasizing the fact that the different questions asked by the two traditions may have different answers" (p. xi). Skyrms's (1996) book articulates the environmental and population parameters under which "rational choice theory completely parts ways with evolutionary theory" (p. 106; see also Cooper 1989). Cognitive mechanisms that were fitness enhancing might well thwart our goals as personal agents in an industrial society (see Baron 1998) because the assumption that our cognitive mechanisms are adapted in the evolutionary sense (Pinker 1997) does not entail normative rationality. Thus, situations where evolutionary and normative rationality dissociate might well put the two processing Systems in partial conflict with each other. These conflicts may be rare, but the few occasions on which they occur might be important ones. This is because knowledge-based, technological societies often put a premium on abstraction and decontextualization, and they sometimes require that the fundamental computational bias of human cognition be overridden by System 2 processes.

6.2. The fundamental computational bias and task interpretation

The fundamental computational bias, that "specific features of problem content, and their semantic associations, constitute the dominant influence on thought" (Evans et al. 1983, p. 295; Stanovich 1999), is no doubt rational in the evolutionary sense. Selection pressure was probably in the direction of radical contextualization. An organism that could bring more relevant information to bear (not forgetting the frame problem) on the puzzles of life probably dealt with the world better than competitors and thus reproduced with greater frequency and contributed more of its genes to future generations.

Evans and Over (1996) argue that an overemphasis on normative rationality has led us to overlook the adaptiveness of contextualization and the nonoptimality of always decoupling prior beliefs from problem situations ("beliefs that have served us well are not lightly to be abandoned," p. 114). Their argument here parallels the reasons that philosophy of science has moved beyond naive falsificationism (see Howson & Urbach 1993). Scientists do not abandon a richly confirmed and well integrated theory at the first little bit of falsifying evidence, because abandoning the theory might actually decrease explanatory coherence (Thagard 1992). Similarly, Evans and Over (1996) argue that beliefs that have served us well in the past should be hard to dislodge, and projecting them on to new information – because of their past efficacy – might actually help in assimilating the new information.

Evans and Over (1996) note the mundane but telling fact that when scanning a room for a particular shape, our visual systems register color as well. They argue that we do not impute irrationality to our visual systems because they fail to screen out the information that is not focal. Our systems of recruiting prior knowledge and contextual information to solve problems with formal solutions are probably likewise adaptive in the evolutionary sense. However, Evans and Over (1996) do note that there is an important disanalogy here as well, because studies of belief bias in syllogistic reasoning have shown that "subjects can to some extent ignore belief and reason from a limited number of assumptions when instructed to do so" (p. 117). That is, in the case of reasoning – as opposed to the visual domain – some people do have the cognitive flexibility to decouple unneeded systems of knowledge and some do not.

The studies reviewed here indicate that those who do have the requisite flexibility are somewhat higher in cognitive ability and in actively open-minded thinking (see Stanovich & West 1997). These styles and skills are largely System 2, not System 1, processes. Thus, the heuristics triggering alternative task construals in the various problems considered here may well be the adaptive evolutionary products embodied in System 1 as Levinson (1995) and others argue. Nevertheless, many of our personal goals may have become detached from their evolutionary context (see Barkow 1989). As Morton (1997) aptly puts it: "We can and do find ways to benefit from the pleasures that our genes have arranged for us without doing anything to help the genes themselves. Contraception is probably the most obvious example, but there are many others. Our genes want us to be able to reason, but they have no interest in our enjoying chess" (p. 106).

Thus, we seek "not evolution's end of reproductive success but evolution's means, love-making. The point of this example is that some human psychological traits may, at least in our current environment, be fitness-reducing" (see Barkow 1989, p. 296). And if the latter are pleasurable, analytic intelligence achieves normative rationality by pursuing *them* – not the adaptive goals of our genes. This is what Larrick et al. (1993) argue when they speak of analytic intelligence as "the set of psychological properties that enables a person to achieve his or her goals effectively. On this view, intelligent people will be more likely to use rules of choice that are effective in reaching their goals than will less intelligent people" (p. 345).

Thus, high analytic intelligence may lead to task construals that track normative rationality; whereas the alternative construals of subjects low in analytic intelligence (and hence more dominated by System 1 processing) might be more likely to track evolutionary rationality in situations that put the two types of rationality in conflict – as is conjectured to be the case with the problems discussed previously. If construals consistent with normative rationality are more likely to satisfy our current individual goals (Baron 1993; 1994) than are construals determined by evolutionary rationality (which are construals determined by our *genes'* metaphorical goal – reproductive success), then it is in this very restricted sense that individual difference relationships such as those illustrated here tell us which construals are "best."

6.3. The fundamental computational bias and the ecology of the modern world

A conflict between the decontextualizing requirements of normative rationality and the fundamental computational bias may perhaps be one of the main reasons that normative and evolutionary rationality dissociate. The fundamental computational bias is meant to be a global term that captures the pervasive bias toward the contextualization of all informational encounters. It conjoins the following processing tendencies: (1) the tendency to adhere to Gricean conversational principles even in situations that lack many conversational features (Adler 1984; Hilton 1995); (2) the tendency to contextualize a problem with as much prior knowledge as is easily accessible, even when the problem is formal and the only solution is a content-free rule (Evans 1982; 1989; Evans et al. 1983); (3) the tendency to see design and pattern in situations that are either undesigned, unpatterned, or random (Levinson 1995); (4) the tendency to reason enthymematically – to make assumptions not stated in a problem and then reason from those assumptions (Henle 1962; Rescher 1988); (5) the tendency toward a narrative mode of thought (Bruner 1986; 1990). All of these properties conjoined represent a cognitive tendency toward radical contextualization. The bias is termed fundamental because it is thought to stem largely from System 1 and that system is assumed to be primary in that it permeates virtually all of our thinking (e.g., Evans & Over 1996). If the properties of this system are not to be the dominant factors in our thinking, then they must be overridden by System 2 processes so that the latter can carry out one of their important functions of abstracting complex situations into canonical representations that are stripped of context. Thus, it is likely that one computational task of System 2 is to decouple (see Navon 1989a; 1989b) contextual features automatically supplied by System 1 when they are potentially interfering.

In short, one of the functions of System 2 is to serve as an override system (see Pollock 1991) for some of the automatic and obligatory computational results provided by System 1. This override function might only be needed in a tiny minority of information processing situations (in most cases, the two Systems will interact in concert), but they may be unusually important ones. For example, numerous theorists have warned about a possible mismatch between the fundamental computational bias and the processing requirements of many tasks in a technological society containing many symbolic artifacts and often requiring skills of abstraction (Adler 1984; 1991; Donaldson 1978; 1993). Hilton (1995) warns that the default assumption that Gricean conversational principles are operative may be wrong for many technical settings because

> many reasoning heuristics may have evolved because they are adaptive in contexts of social interaction. For example, the expectation that errors of interpretation will be quickly repaired may be correct when we are interacting with a human being but incorrect when managing a complex system such as an aircraft, a nuclear power plant, or an economy. The evolutionary adaptiveness of such an expectation to a conversational setting may explain why people are so bad at dealing with lagged feedback in other settings. (p. 267)

Concerns about the real-world implications of the failure to engage in necessary cognitive abstraction (see Adler 1984) were what led Luria (1976) to warn against minimizing the importance of decontextualizing thinking styles. In discussing the syllogism, he notes that "a considerable proportion of our intellectual operations involve such verbal and logical systems; they comprise the basic network of codes along which the connections in discursive human thought are channeled" (p. 101). Likewise, regarding the subtle distinctions on many decontextualized language tasks, Olson (1986) has argued that "the distinctions on which such questions are based are extremely important to many forms of intellectual activity in a literate society. It is easy to show that sensitivity to the subtleties of language are crucial to some undertakings. A person who does not clearly see the difference between an expression of intention and a promise or between a mistake and an accident, or between a falsehood and a lie, should avoid a legal career or, for that matter, a theological one" (p. 341). Objective measures of the requirements for cognitive abstraction have been increasing across most job categories in technological societies throughout the past several decades (Gottfredson 1997). This is why measures of the ability to deal with abstraction remain the best employment predictor and the best earnings predictor in postindustrial societies (Brody 1997; Gottfredson 1997; Hunt 1995).

Einhorn and Hogarth (1981) highlighted the importance of decontextualized environments in their discussion of the optimistic (Panglossian/Apologist) and pessimistic (Meliorist) views of the cognitive biases revealed in laboratory experimentation. They noted that "the most optimistic asserts that biases are limited to laboratory situations which are unrepresentative of the natural ecology" (p. 82), but they go on to caution that "in a rapidly changing world it is unclear what the relevant natural ecology will be. Thus, although the laboratory may be an unfamiliar environment, lack of ability to perform well in unfamiliar situations takes on added importance" (p. 82). There is a caution in this comment for critics of the abstract content of most laboratory tasks and standardized tests. The issue is that, ironically, the argument that the laboratory tasks and tests are not like "real life" is becoming less and less true. "Life," in fact, is becoming more like the tests!

The cognitive ecologists have, nevertheless, contributed greatly in the area of remediation methods for our cognitive deficiencies (Brase et al. 1998; Cosmides & Tooby 1996; Fiedler 1988; Gigerenzer & Hoffrage 1995; Sedlmeier 1997). Their approach is, however, somewhat different from that of the Meliorists. The ecologists concentrate on shaping the environment (changing the stimuli presented to subjects) so that the same evolutionarily adapted mechanisms that fail the standard of normative rationality under one framing of the problem give the normative response under an alternative (e.g., frequentistic) version. Their emphasis on environmental alteration provides a much-needed counterpoint to the Meliorist emphasis on cognitive change. The latter, with their emphasis on reforming human thinking, no doubt miss opportunities to shape the environment so that it fits the representations that our brains are best evolved to deal with. Investigators framing cognition within a Meliorist perspective are often blind to the fact that there may be remarkably efficient mechanisms available in the brain – if only it was provided with the right type of representation.

On the other hand, it is not always the case that the world will let us deal with representations that are optimally suited to our evolutionarily designed cognitive mechanisms. For example, in a series of elegant experiments, Gigerenzer et al. (1991) have shown how at least part of the overconfidence effect in knowledge calibration studies is due to the unrepresentative stimuli used in such experiments – stimuli that do not match the subjects' stored cue validities, which are optimally tuned to the environment. But there are many instances in real-life when we are suddenly placed in environments where the cue validities have changed. Metacognitive awareness of such situations (a System 2 activity) and strategies for suppressing incorrect confidence judgments generated by the responses to cues automatically generated by System 1 will be crucial here. High school musicians who aspire to a career in music have to recalibrate when they arrive at university and encounter large numbers of talented musicians for the first time. If they persist in their old confidence judgments they may not change majors when they should. Many real-life situations where accomplishment yields a new environment with even more stringent performance requirements share this logic. Each time we "ratchet up" in the competitive environment of a capitalist economy we are in a situation just like the overconfidence knowledge calibration experiments with their unrepresentative materials (Frank & Cook 1995). It is important to have learned System 2 strategies that will temper one's overconfidence in such situations (Koriat et al. 1980).

7. Individual differences and the normative/descriptive gap

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 the cause of the gap between normative models and descriptive models of actual performance. Patterns of individual differences have implications for arguments that all such gaps reflect merely performance errors. Individual differences are also directly relevant to theories that algorithmic-level limitations prevent the computation of the normative response in a system that would otherwise compute it. The wrong norm and alternative construal explanations of the gap involve many additional complications but, at the very least, patterns of individual differences might serve as "intuition pumps" (Dennett 1980) and alter our reflective equilibrium regarding the plausibility of such explanations (Stanovich 1999).

Different outcomes occurred across the wide range of tasks we have examined in our research program. Of course, all the tasks had some unreliable variance and thus some responses that deviated from the response considered normative could easily be considered as performance errors. But not all deviations could be so explained. Several tasks (e.g., syllogistic reasoning with interfering content, four-card selection task) were characterized by heavy computational loads that made the normative response not prescriptive for some subjects – but these were usually few in number.[13] Finally, a few tasks yielded patterns of covariance that served to raise doubts about the appropriateness of normative models applied to them and/or the task construals assumed by the problem inventors (e.g., several noncausal baserate items, false consensus effect).

Although many normative/descriptive gaps could be reduced by these mechanisms, not all of the discrepancies could be explained by factors that do not bring human rationality into question. Algorithmic-level limitations were far from absolute. The magnitude of the associations with cognitive ability left much room for the possibility that the remaining reliable variance might indicate that there are systematic irrationalities in intentional-level psychology. A component of our research program mentioned only briefly previously has produced data consistent with this possibility. Specifically, it was not the case that once capacity limitations had been controlled the remaining variations from normative responding were unpredictable (which would have indicated that the residual variance consisted largely of performance errors). In several studies, we have shown that there was significant covariance among the scores from a variety of tasks in the heuristics and biases literature after they had been residualized on measures of cognitive ability (Stanovich 1999). The residual variance (after partialling cognitive ability) was also systematically associated with questionnaire responses that were conceptualized as intentional-level styles relating to epistemic regulation (Sá et al. 1999; Stanovich & West 1997; 1998c). Both of these findings are indications that the residual variance is systematic. They falsify models that attempt to explain the normative/descriptive gap entirely in terms of computational limitations and random performance errors. Instead, the findings support the notion that the normative/descriptive discrepancies that remain after computational limitations have been accounted for reflect a systematically suboptimal intentional-level psychology.

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. For example, 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). But from a psychological standpoint, there may be important implications in precisely the aspects of performance that have been backgrounded in this controversy ("adventitious causes"). For example, Johnson-Laird and Byrne (1993) articulate 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 systematically varying cognitive styles to play a more important role in theories of rational thought. At the heart of the rational competence that Johnson-Laird and Byrne (1993) attribute to humans is not perfect rationality but instead just one meta-principle: 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" (p. 178).

The key point in Johnson-Laird and Byrne's (1991; 1993; see Johnson-Laird 1999) account[14] is that once an individual constructs a mental model from the premises, once the individual draws a new conclusion from the model, and once the individual begins the search for an alternative model of the premises which contradicts the conclusion, the individual "lacks any systematic method to make this search for counter-examples" (p. 205; see Bucciarelli & Johnson-Laird 1999). Here is where Johnson-Laird and Byrne's (1993) model could be modified to allow for the influence of thinking styles in ways that the impeccable competence view of Cohen (1981; 1982) does not. In this passage, Johnson-Laird and Byrne seem to be arguing that there are no systematic control features of the search process. But styles of epistemic regulation (Sá et al. 1999; Stanovich & West 1997) 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" (see Johnson-Laird & Oatley 1992; Oatley 1992; Overton 1985; 1990) – factors such as dispositions toward premature closure, cognitive confidence, reflectivity, dispositions toward confirmation bias, ideational generativity, and so on.

Dennett (1988) argues that we use the intentional stance for humans and dogs but not for lecterns because for the latter "there is no predictive leverage gained by adopting the intentional stance" (p. 496). In the experiments just mentioned (Sá et al. 1999; Stanovich & West 1997; 1998c), it has been shown that there is additional predictive leverage to be gained by relaxing the idealized rationality assumption of Dennett's (1987; 1988) intentional stance and by positing measurable and systematic variation in intentional-level psychologies. Knowledge about such individual differences in people's intentional-level psychologies can be used to predict variance in the normative/descriptive gap displayed on many reasoning tasks. Consistent with the Meliorist conclusion that there can be individual differences in human rationality, our results show that there is variability in reasoning that cannot be accommodated within a model of perfect rational competence operating in the presence of performance errors and computational limitations.

NOTES

1. Individual differences on tasks in the heuristics and biases literature have been examined previously by investigators such as Hoch and Tschirgi (1985), Jepson et al. (1983), Rips and Conrad (1983), Slugoski and Wilson (1998), and Yates et al. (1996). Our focus here is the examination of individual differences through a particular metatheoretical lens – as providing principled constraints on alternative explanations for the normative/descriptive gap.

2. All of the work cited here was conducted within Western cultures which matched the context of the tests. Of course, we recognize the inapplicability of such measures as indicators of cognitive ability in cultures other than those within which the tests were derived (Ceci 1996; Greenfield 1997; Scribner & Cole 1981). Nevertheless, it is conceded by even those supporting more contextualist views of intelligence (e.g., Sternberg 1985; Sternberg & Gardner 1982) that measures of general intelligence do identify individuals with superior reasoning ability – reasoning ability that is then applied to problems that may have a good degree of cultural specificity (see Sternberg 1997; Sternberg & Kaufman 1998).

3. The Scholastic Assessment Test is a three-hour paper-and-pencil exam used for university admissions testing. The verbal section of the SAT test includes three types of items: verbal analogies, sentence completions, and critical reading problems. The mathematical section contains arithmetic, algebra, and geometry problems that require quantitative reasoning.

4. We note that the practice of analyzing a single score from such ability measures does not imply the denial of the existence of second-order factors in a hierarchical model of intelligence. However, theorists from a variety of persuasions (Carroll 1993; 1997; Hunt 1997; Snyderman & Rothman 1990; Sternberg & Gardner 1982; Sternberg & Kaufman 1998) acknowledge that the second-order factors are correlated. Thus, such second-order factors are not properly interpreted as separate faculties (despite the popularity of such colloquial interpretations of so-called "multiple intelligences"). In the most comprehensive survey of intelligence researchers, Snyderman and Rothman (1990) found that by a margin of 58% to 13%, the surveyed experts endorsed a model of "a general intelligence factor with subsidiary group factors" over a "separate faculties" model. Throughout this target article we utilize a single score that loads highly on the general factor, but analyses which separated out group factors (Stratum II in Carroll's widely accepted model based on his analysis of 460 data sets; see Carroll 1993) would reveal convergent trends.

5. Positive correlations with developmental maturity (e.g., Byrnes & Overton 1986; Jacobs & Potenza 1991; Klahr et al. 1993; Markovits & Vachon 1989; Moshman & Franks 1986) would seem to have the same implication.

6. However, we have found (Stanovich & West 1999) that the patterns of individual differences reversed somewhat when the potentially confusing term "false positive rate" was removed from the problem (see Cosmides & Tooby 1996 for work on the effect of this factor). It is thus possible that this term was contributing to an incorrect construal of the problem (see sect. 5).

7. However, sometimes alternative construals might be computational escape hatches (Stanovich 1999). That is, an alternative construal might be hiding an inability to compute the normative model. Thus, for example, in the selection task, perhaps some people represent the task as an inductive problem of optimal data sampling in the manner that Oaksford and Chater (1994; 1996) have outlined because of the difficulty of solving the problem if interpreted deductively. As O'Brien (1995) demonstrates, the abstract selection task is a very hard problem for a mental logic without direct access to the truth table for the material conditional. Likewise, Johnson-Laird and Byrne (1991) have shown that tasks requiring the generation of counterexamples are difficult unless the subject is primed to do so.

8. The results with respect to the framing problems studied by Frisch (1993) do not always go in this direction. See Stanovich and West (1998b) for examples of framing problems where the more cognitively able subjects are not less likely to display framing effects.

9. Kahneman and Tversky (1982) themselves (pp. 132–35) were among the first to discuss the issue of conversational implicatures in the tasks employed in the heuristics and biases research program.

10. Of course, another way that cognitive ability differences might be observed is if the task engages only System 2. For the present discussion, this is an uninteresting case.

11. It should be noted that the distinction between normative and evolutionary rationality used here is different from the distinction between rationality$_1$ and rationality$_2$ used by Evans and Over (1996). They define rationality$_1$ as reasoning and acting "in a way that is generally reliable and efficient for achieving one's goals" (p. 8). Rationality$_2$ concerns reasoning and acting "when one has a reason for what one does sanctioned by a normative theory" (p. 8). Because normative theories concern goals at the personal level, not the genetic level, both of the rationalities defined by Evans and Over (1996) fall within what has been termed here normative rationality. Both concern goals at the personal level. Evans and Over (1996) wish to distinguish the explicit (i.e., conscious) following of a normative rule (rationality$_2$) from the largely unconscious processes "that do much to help them achieve their ordinary goals" (p. 9). Their distinction is between two sets of algorithmic mechanisms that can both serve normative rationality. The distinction we draw is in terms of *levels* of optimization (at the level of the replicator itself – the gene – or the level of the vehicle); whereas theirs is in terms of the *mechanism* used to pursue personal goals (mechanisms of conscious, reason-based rule following versus tacit heuristics).

It should also be noted that, for the purposes of our discussion here, the term evolutionary rationality has less confusing connotations than the term "adaptive rationality" discussed by Oaksford and Chater (1998). The latter could potentially blur precisely the distinction stressed here – that between behavior resulting from adaptations in service of the genes and behavior serving the organism's current goals.

12. Evidence for this assumption comes from voluminous data indicating that analytic intelligence is related to the very type of outcomes that normative rationality would be expected to maximize. For example, the System 2 processes that collectively comprise the construct of cognitive ability are moderately and reliably correlated with job success and with the avoidance of harmful behaviors (Brody 1997; Lubinski & Humphreys 1997; Gottfredson 1997).

13. Even on tasks with clear computational limitations, some subjects from the lowest strata of cognitive ability solved the problem. Conversely, on virtually all the problems, some university subjects of the highest cognitive ability failed to give the normative response. Fully 55.6% of the university subjects who were at the 75th percentile or above in our sample in cognitive ability committed the conjunction fallacy on the Linda Problem. Fully 82.4% of the same group failed to solve a nondeontic selection task problem.

14. A reviewer has pointed out that the discussion here is not necessarily tied to the mental models approach. The notion of searching for counterexamples under the guidance of some sort of control process is at the core of any implementation of logic.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Three fallacies

Jonathan E. Adler
Department of Philosophy, Brooklyn College and the Graduate Center of the City University of New York, Brooklyn, NY 11210
jadler@brooklyn.cuny.edu

Abstract: Three fallacies in the rationality debate obscure the possibility for reconciling the opposed camps. I focus on how these fallacies arise in the view that subjects interpret their task differently from the experimenters (owing to the influence of conversational expectations). The themes are: first, critical assessment must start from subjects' understanding; second, a modal fallacy; and third, fallacies of distribution.

Three fallacies in the rationality debate obscure the possibility for reconciling the opposed camps, a reconciliation toward which Stanovich & West (S&W) display sympathy in their discussion of dual models and the understanding/accepting principle.

The fallacies are prominent in S&W's treatment of the response they take as most challenging: "subjects have a different interpretation of the task" (sect. 5, para. 1). The response requires the *subjects' understanding* assumption: "criticism of subjects' performance must start from the subjects' understanding." The argument then is that if subjects' responses are correct according to their own (reasonable) interpretation of their task, then they are correct (Adler 1994).

But consider conversational reinterpretations of Piaget's experiments on conservation. Children younger than 6-years-of-age deny that the length of sticks remains the same after rearrangement, and that is explained as satisfying the expectation of *relevance* of the experimenter's action (in moving the objects). However, if a child over 7 offered the nonconserving response, we would regard that as a defect in reasoning, even though it was perfectly in accord with his (reasonable) interpretation. This child has not achieved the common knowledge that not all activities by speakers (or actors) along with their focal contribution are maximally relevant to it.

Contrary to the subjects' understanding assumption, such contextualized interpretations may reflect a defective or weak grasp of the crucial logical terms and formal principles that the experimenters are studying. The weak grasp helps explain why subjects so readily read their task differently from the experimenters. Consider the base-rate studies. It is plausible that understanding "probability" as something close to "explanatorily coherent" or "reasonable, based on the evidence" fits ordinary usage and reasoning better. Although that interpretation may be suggested and even encouraged, it is not mandated. (Gricean implicatures are *required,* not merely allowed or suggested; Grice 1989) The interpretations of their task that subjects must make to accord with explanations like the conversational one depend upon their failure to have accessible or online grasp of important distinctions.

Compare this to teaching: I understand why students read "Jones does not believe that Marcia is in Italy" as "Jones believes that Marcia is not in Italy." Roughly, negation is understood (conversationally) to function as a denial, and a denial is typically informative only if treated as having small-scope reading. But it is my job to show students that there is an important distinction there that their understanding obscures and that encourages faulty inference. I expose the defects through eliciting other, related judgments of theirs. The failing is a failing according to their own beliefs.

The fallacy in the subjects'-understanding assumption is an instance of the broader one in the saying "to understand is to forgive." To understand subjects' responses conversationally need not be to justify them.

To infer from the correctness of subjects' judgments given their own construal of their task the correctness of subjects' judgments is to commit a second (modal) fallacy:

> If *S* understands problem *P* as *t* and if answer *A* follows from (is the right answer to) *t,* then *S* reasons well when *S* answers *A* to problem P.
> *S* understands problem P as *t* and answer *A* does follow from *t.*
> So, *S* reasons well when *S* answers *A* to *t.*

The conclusion does not follow, as it does not in the following:

> If John is a bachelor, then he must be unmarried.
> John is a bachelor.
> So, John must be unmarried.

The fallacy arises from the natural misplacement of the "must" in the consequent clause. The necessity really governs the whole statement. Similarly, if we place the "reasons well," from the previous argument sketch, out front, it is evident that the conclusion does not follow, unless we also know that:

> *S* reasons well in understanding *P* as *t.*

But children do not reason well in the Piaget task by understanding the question actually asked "Are the sticks the same size or is one longer?" as having the meaning "Which answer – that the sticks are the same size or that one is longer – renders the experimenter's actions (more) relevant to his actual question?" (Compare to Margolis 1987: Ch. 8.)

The third and last fallacy is a distributional one. Consider, specifically, the base rate studies applied to judgments of one's marital prospects, where the divorce rates are to serve as base rates. Gilovich (1991) comments: "To be sure, we should not discount our current feelings and self-knowledge altogether; we just need to temper them a bit more with our knowledge of what happens to people in general. This is the consensus opinion of all scholars in the field."

But if belief is connected to action in the standard ways, the recommendation does not accord with the marriage *practice,* since if a prospective spouse lessens her commitment, she cannot expect her partner not to do so in turn, threatening an unstable downward spiral. But it is equally wrong to conclude that one should never temper one's judgments, nor that for purposes of prediction or betting, one should not integrate the divorce rates with one's self-knowledge.

But let us revert to our main illustration by considering Gilovich's "consensus" recommendation as applied to conversation. Imagine that we acceded to the recommendation to temper our acceptance of testimony by an estimate of the degree of reliable truth telling in the relevant community. The result would be that much more of what we standardly accept as true we can now only accept as bearing some high degree of probability. Our tempered beliefs issue in tempered acceptance. The layers of complexity are unmanageable. We will undermine the trust and normal flow of information between speakers and hearers. No conversational practice corresponds to it.

The (distributional) fallacy in this (Meliorist) direction is to infer that because in any particular case one can so temper one's acceptance of the word of another, one can do so regularly. But the opposed (Panglossian) fallacy is to infer that because the practice forbids tempering generally, in each case one is justified in not tempering. This opposite distributional fallacy is a form of rule-worship – it is to infer that we ought never to make an exception in accord with the details of the case. As above, it is right for us to expect relevance, but it is wrong not to allow that expectation to be overruled.

Do the birds and bees need cognitive reform?

Peter Ayton
Department of Psychology, City University, London EC1V 0HB, United Kingdom **p.ayton@city.ac.uk**

Abstract: Stanovich & West argue that their observed positive correlations between performance of reasoning tasks and intelligence strengthen the standing of normative rules for determining rationality. I question this argument. Violations of normative rules by cognitively humble creatures in their natural environments are more of a problem for normative rules than for the creatures.

What is "normative"? The assumption that decisions can be evaluated in relation to uncontentious abstract rules has proved alluring. More than 300 years ago Leibniz envisaged a universal calculus or characteristic by means of which all reasonable people would reach the same conclusions. No more disputes:

> "Its authority will not be open any longer to doubt when it becomes possible to reveal the reason in all things with the clarity and certainty which was hitherto possible only in arithmetic. It will put an end to that sort of tedious objecting with which people plague each other, and which takes away the pleasure of reasoning and arguing in general. (Leibniz 1677/1951, p. 23)

His plan was simple: characteristic numbers would be established for *all* ideas and all disputes would be reduced to computations involving those numbers. Leibniz felt that establishing the characteristic numbers for all ideas would not take all that long to implement:

> I believe that a few selected persons might be able to do the whole thing in five years, and that they will in any case after only two years arrive at a mastery of the doctrines most needed in practical life, namely, the propositions of morals and metaphysics, according to an infallible method of calculation." (pp. 22–23)

What became of this plan? Though Leibniz was not alone, before long mathematicians and philosophers gave up the task of reducing rationality to a calculus (see Daston 1988). Curiously, psychologists have not (see Gigerenzer 1996b). The Stanovich & West (S&W) target article claims that the systematic association of normative responses with intelligence is embarrassing to those who feel that norms are being incorrectly applied. But, as the authors admit, there are unembarrassing accounts for this phenomenon: where there is ambiguity about how to construe the task, people who are educated or motivated to reason with mathematics or logic might do so, while others reason according to some other scheme.

In their experiments on the Linda problem, Hertwig and Gigerenzer (1999) showed that people impute a nonmathematical meaning to the term "probability"; their reasonable pragmatic inferences then result in violations of the conjunction rule. Nonetheless, in different contexts, subjects correctly interpreted the mathematical meaning of the term "probability" and were less likely to violate the rule. Sensitivity to the cues for different interpretations is likely to be conditioned by subjects' education, motivation, and cognitive ability. This hypothesis could be tested. However, if more intelligent people used the mathematical interpretation more often, why should this condemn the responses of those who reasoned differently?

S&W argue that we should try to avoid the conclusion that those with more computational power are systematically computing the nonnormative response and that, in any case, such an outcome would be "an absolute first." However, by the narrow standards of normative theory, there is evidence of exactly this phenomenon. Ayton and Arkes (1998) and Arkes and Ayton (1999) review studies that show that adults commit an error contrary to the normative cost-benefit rules of choice, whereas children and even cognitively humble nonhuman organisms do not. Yet one need not – unless one is committed to judging by the normative rule alone – conclude that adults are in any general sense less rational. According to the notion of bounded rationality (Simon 1956; 1992), both the computational limits of cognition and the structure of the environment may foster the use of "satisficing" strategies that are effective despite violating normative rules.

The inadequacy of judging by normative rules is brought into focus when we contemplate how we would deal with evidence of normative violations in cognitively humble lower animals' decisions: would it make sense to claim that they are irrational? For example, transitivity of choice is considered one of the cornerstones of classical rationality (Fishburn 1991; McClennan 1990). Nonetheless, Shafir (1994b) has shown that honey bees violate transitivity in choosing between flowers. Should we conclude that they are irrational and in need of cognitive reform? Bees have been successfully going about their business for millions of years. As they have managed to survive for as long as they have, whilst violating one of the basic axioms of rationality, one feels that it is the axiom that is limited in capturing what it means to be rational – not the bees.

Shafir explained that the intransitivities indicate that bees make comparative rather than absolute evaluations of flowers. Tversky (1969) suggested that comparative decision-making was more efficient than absolute evaluation – it requires less processing. So, once one takes their environment and the resource constraints into account it may well be that bees' behaviour is optimal – despite not being predictable from any normative model.

Other researchers claim that wasps, birds, and fish commit the sunk cost fallacy (Ayton & Arkes 1998; Arkes & Ayton 1999). In wild animals, natural selection will ruthlessly expunge any strategy that can be bettered at no extra cost but, of course, in nature there are resource constraints. The extra computational resources needed to behave as normative theory requires might be prohibitive – given that "fast and frugal" strategies operating in natural environments can be highly effective whilst clearly violating axioms of rationality (Gigerenzer & Goldstein 1996). Birds do it, bees do it, even educated Ph.D.s do it; why not violate normative rules of rationality?

Slovic and Tversky's (1974) paper beautifully illustrates the vulnerability of normative theory in its dependence on the acceptance of unverifiable axioms; they did *not* demonstrate that the deeper the understanding of the independence axiom, the greater the readiness to accept it – but even if they had this would hardly "prove" it. Given his rejection of the axiom, should we assume that Allais (1953) doesn't understand it? Feeling uncertain, Edwards (1982) polled decision theorists at a conference. They unanimously endorsed traditional Subjective Expected Utility theory as the appropriate normative model and unanimously agreed that people do not behave as that model requires. S&W's interpretation of the positive manifold might be seen as an election about normative rules – where votes are weighted according to IQ. But science does not progress through elections.

Finally, two terminological concerns. "Meliorist" is an odd term for those who assume that irrationality results from the inherent nature of cognition. The term "Panglossian" is somewhat ironic given that the inspiration for Voltaire's comic character Dr. Pangloss was that dreamer of the power of logic, Leibniz.

Alternative task construals, computational escape hatches, and dual-system theories of reasoning

Linden J. Ball and Jeremy D. Quayle

Institute of Behavioural Sciences, University of Derby, Mickleover, Derby, DE3 5GX, United Kingdom **{l.j.ball; j.d.quayle}@derby.ac.uk**
IBS.derby.ac.uk/staff/Linden_Ball.html

Abstract: Stanovich & West's dual-system represents a major development in an understanding of reasoning and rationality. Their notion of System 1 functioning as a computational escape hatch during the processing of complex tasks may deserve a more central role in explanations of reasoning performance. We describe examples of apparent escape-hatch processing from the reasoning and judgement literature.

Stanovich & West (S&W) present impressive support for their proposal that patterns of individual differences in performance can advance an understanding of reasoning, rationality, and the normative/descriptive gap. We find their evidence and arguments compelling, and likewise believe that dual-system accounts are central to clarifying the nature and limits of human rationality. Many of S&W's proposals surrounding the goals, constraints, and operations of System 1 (contextualised, interactional intelligence) and System 2 (decontextualised, analytic intelligence) strike us as significant conceptual advances over previous dual-process accounts (Epstein 1994; Evans & Over 1996; Sloman 1996), which, because of elements of under-specification, have often raised as many questions as they have answered. Other strengths of S&W's account derive from their recognition of the importance of intentional-level constructs (e.g., metacognition and thinking styles) in controlling cognition.

Still, certain claims about how a dual-system distinction can explain performance dichotomies between individuals of differing analytic intelligence seem worthy of critical analysis. One claim that forms our focus here is that "sometimes alternative construals [arising from System 1] might be computational escape hatches. . . . That is, an alternative construal might be hiding an inability to compute the normative model" (S&W, n. 7). As an example S&W note that some people may process abstract selection tasks as inductive problems of optimal data sampling (Oaksford & Chater 1994) because of difficulties in computing deductive responses via System 2. This computational-escape-hatch concept is appealing; we have alluded to a similar notion (Ball et al. 1997, p. 60) when considering the processing demands (e.g., relating to the requirement for meta-inference) of abstract selection tasks. We wonder, however, whether the computational-escape-hatch idea should feature more centrally in S&W's dual-system account so that it may generalise findings across a range of difficult tasks.

To explore this possibility it is necessary to examine S&W's proposals regarding the application of analytic abilities to override System 1 task construals. They state that "for some individuals, System 2 processes operating in parallel . . . will have the requisite computational power . . . to override the response primed by System 1" (sect. 6.1, para. 5), and further note that this override function "might only be needed in a tiny minority of information processing situations (in most cases, the two Systems will interact in concert)" (sect. 6.3, para. 2). What we find revealing here is the suggestion that all individuals will at least attempt to apply System

2 processes to achieve some form of decontextualised task construal – albeit perhaps a fragmentary one. Having put the effort into System 2 computation it is hard to imagine why any individual (even ones low in analytic intelligence) should then ignore the System 2 output, unless, perhaps, they lack confidence about the efficacy of their System 2 computations (i.e., they have metacognitive awareness of having experienced computational difficulties). Indeed considerable evidence exists that people do produce normatively-optimal responses to computationally-tractable deductive-reasoning problems (e.g., certain "one-model" syllogisms) and that common non-normative responses to harder problems reflect possible (but not necessary) inferences from attempts at applying a deductive procedure (e.g., Johnson-Laird & Byrne 1991). Perhaps, then, tasks at the easy-to-intermediate end of the complexity continuum do invoke System 2 responses for most individuals, whereas tasks at the intermediate-to-hard end differentially favour a System 2 response from those of higher analytic intelligence, and a last-resort System 1 response from those of lower analytic intelligence (because they induce high levels of metacognitive uncertainty about System 2 efficacy).

One upshot of this argument is that recourse to System 1 computational escape hatches may, for any individual, vary from problem to problem depending on processing demands and levels of metacognitive uncertainty about System 2 functioning. Thus, whilst performing on nondeontic selection tasks may, for the majority, reflect either a fall-back System 1 response or a failed attempt at System 2 processing, performance on deontic versions may be within nearly everyone's System 2 capabilities. Indeed, Johnson-Laird (1995) presents an essentially System 2 account of why deontic selections tasks (and nondeontic ones where counterexamples can be invoked) may be easy to compute normative responses for. If deontic selection tasks reflect manageable System 2 processing, then this obviates any need to posit System 1 task construals (e.g., based around pragmatic schemas or Darwinian algorithms).

Another upshot of our argument about task difficulty, metacognitive uncertainty and fall-back mechanisms is the possibility that an escape-hatch response may actually be the default strategy for any individual whose motivated attempt at a System 2 construal is overloaded. As such, computational escape hatches may underlie more responding than S&W seem willing to concede. One example from the judgement literature is Pelham et al.'s (1994) proposal that people fall back on a "numerosity = quantity" heuristic when judging amount (e.g., of food) under conditions of task complexity. Another example comes from our own account on belief-bias effects in the evaluation of syllogistic conclusions (Quayle & Ball, in press) which assumes that participants: (1) fulfil instructions to suspend disbelief and accept the truth of implausible premises (i.e., by overriding initial System 1 processing); (2) attempt the (System 2) application of a mental-models based reasoning strategy; and (3) produce belief-biased responses (i.e., use System 1 as an escape hatch) when working-memory constraints lead to threshold levels of metacognitive uncertainty being surpassed in relation to the perceived efficacy of (System 2) reasoning. The latter, we argue, is more likely to happen on invalid than valid syllogisms since invalid syllogisms are difficult (see Ball & Quayle 1999; Hardman & Payne 1995, for supporting evidence), so explaining the standard interaction between belief and logic on conclusion acceptances. This escape-hatch account predicts that participants will be more confident with responses to valid than invalid problems, and more belief-biased with invalid problems when they have lower working-memory capacities than fellow reasoners. Our data support both predictions (Quayle & Ball, in press) and are difficult to reconcile with theories positing selective scrutiny of unbelievable conclusions (e.g., Evans et al. 1993).

In conclusion, although we believe that S&W's proposals are a milestone in the development of an understanding of reasoning and rationality, we feel they may have downplayed the role of System 1 functioning as a computational escape hatch (whether triggered by algorithmic-level limitations or intentional-level factors). To test predictions of escape-hatch accounts of reasoning would seem a fruitful avenue for investigation using process-tracing methods, including protocol analysis, eye tracking and on-line confidence assessment. Such techniques should help clarify aspects of metacognitive processing and the flow of control between dual systems during reasoning.

Normative and prescriptive implications of individual differences

Jonathan Baron

Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6196 **baron@psych.upenn.edu**
www.sas.upenn.edu/~jbaron

Abstract: Stanovich & West (S&W) have two goals, one concerned with the evaluation of normative models, the other with development of prescriptive models. Individual differences have no bearing on normative models, which are justified by analysis, not consensus. Individual differences do, however, suggest where it is possible to try to improve human judgments and decisions through education rather than computers.

The extensive research program described in the target article is apparently directed at two goals, one concerned with the evaluation of normative models, the other concerned with exploration of the possibilities for improvement of human judgment and decisions. The latter seems more promising to me.

Stanovich & West (S&W) occasionally imply that normative models gain support when smarter people endorse their conclusions or when arguments on both sides lead people to endorse them (the understanding-acceptance principle). In my view (Baron, in press), normative models should not be justified by consensus, even by consensus of experts or smart people.

Normative models, I argue, come from the analysis of situations. A good analogy is arithmetic. If you put two drops of water into a flask, and then another two, you may get fewer than four units of water in the flask, because the drops join together (Popper 1962). Is this arithmetic falsified? We usually think not, because we carefully define the application of counting so as to exclude the coalescing of drops.

Similarly, expected-utility theory (for example) comes from analysis of certain situations into uncertain states of the world (beyond our control), acts (under our control), and outcomes (what we care about), which depend jointly on the acts and states. We further assume that our caring about outcomes defines a dimension along which differences are meaningful. Like arithmetic, this is an idealization, but a useful one.

Normative models like expected-utility theory provide standards for the evaluation of judgments and decisions. If we define the standards according to the intuitions of a majority of experts, then we can never argue against the majority. And what is normative can change over time.

A good example, which S&W have not studied (yet), is the ambiguity effect. People prefer to bet on a red (or white) ball being drawn from an urn containing 50 red balls and 50 white ones than to bet on a red (or white) ball being drawn from an (ambiguous) urn containing an unknown proportion of red and white balls (even if they pick the color, and even if the prize for the first urn is slightly less). This kind of preference yields responses inconsistent with a normative principle of independence (Baron, in press). This principle can be derived from logic, once decisions have been analyzed into acts, states, and outcomes. It is about as clearly normative as you can get, yet it is so strongly against our intuition that several scholars have simply rejected it on this basis (e.g., Ellsberg 1961; Rawls 1971).

But where does intuition – even expert intuition – get such authority? Must standards of reasoning depend on a leap of faith in its power? I think not. We can, instead, understand our intuitions as misunderstandings, overgeneralizations of principles that usually work. The ambiguity effect results from a principle against choosing options when we lack information about their outcomes. This is irrelevant to the urn problem, because the information is unavailable. We know this, but our intuitive judgment is based on a simpler description, which does not take into account our lack of access to the composition of the ambiguous urn.

In sum, we need not accept the argument that intuitive judgments – whether made by students in experiments or by philosophers – have normative authority. Even if all subjects gave the non-normative answer and all philosophers' intuitions agreed with the subjects' responses, the proposed normative model could still be correct. Likewise, the fact that "experts disagree" is not an answer to an argument for some particular normative model. Truth is defined by arguments and evidence, not consensus.

Although the S&W research program does not bear on the validity of normative models, it is highly relevant to prescriptive questions, that is, questions about who should do what when judgments and decisions fall short of normative standards. It is these prescriptive questions that, in the end, justify our research. The heuristics-and-biases tradition is most valuable as an answer to questions of whether, and how, we can improve decisions in the real world. (If the answer is that we cannot, because people are already doing as well as possible, then this is a pessimistic conclusion, as S&W note.) Laboratory research is limited in what it can tell us about thinking outside the laboratory, but, when we combine laboratory results with observation of real-world problems (e.g., the chapter on risk in Baron, in press), its implications for practice can be as helpful in public policy as biology research can be in medicine. As in medicine, we must test the implications in the "clinic" too.

The results on individual differences suggest that, in some cases, we can hope to teach people to improve their judgment on their own. Some people are already doing well according to normative standards. In many of these cases, it is difficult to see how limitations on mental capacity in the narrow sense (Baron 1985) can prevent others people from improving. In other cases, such as some of the observed failures to use prior probabilities, it seems that few can learn to improve. When we find this, the thing to do is not to throw out the normative model, but, rather, to rely more on external aids, such as computers.

Reasoning strategies in syllogisms: Evidence for performance errors along with computational limitations

Monica Bucciarelli

Centro di Scienze Cognitiva, Università di Torino, 10123 Turin, Italy
monica@psych.unito.it

Abstract: Stanovich & West interpret errors in syllogistic reasoning in terms of computational limitations. I argue that the variety of strategies used by reasoners in solving syllogisms requires us to consider also performance errors. Although reasoners' performance from one trial to another is quite consistent, it can be different, in line with the definition of performance errors. My argument has methodological implications for reasoning theories.

Stanovich & West (S&W) define *performance errors* as algorithm-level problems that are transitory in nature, and *computational limitations* as nontransitory problems at the algorithmic level that would be expected to recur on a readministration of the task. The authors find covariance between ability in the Scholastic Assessment Test and – among others – the ability to draw syllogistic inferences. They conclude that the gap between normative and descriptive syllogistic reasoning can to a moderate extent be accounted for by variation in computational limitations. They claim that positing errors at this level is legitimate.

Alas, the proposal by S&W does not encompass the experimental data on the different performances of reasoners dealing with the same problem on two different trials. As a consequence, S&W tend to underestimate the role of transitory algorithm-level problems in syllogistic reasoning. Indeed, the experimental results of Johnson-Laird and Steedman (1978) show that adults' performance in syllogistic reasoning can vary significantly when the same problem is encountered twice. More recently, in line with these results, Bucciarelli and Johnson-Laird (1999) observed a great variability in performance – both between and within reasoners – in the strategies that they used to draw conclusions from syllogistic premises. Within the framework of the theory of mental models (Johnson-Laird 1983; Johnson-Laird & Byrne 1991), they carried out a series of experiments to observe the sequence of external models that the participants built in drawing their own conclusions from syllogistic premises. These external models took the form of shapes representing the various terms in the syllogisms, which the participants could use help them reason. In one experiment (Experiment 4), each participant carried out the inferential task in two different experimental conditions for purposes of comparison: once using the external models, and once without using them. Their performance is best interpreted in terms of computational limitations and performance errors.

Errors can be interpreted as the result of computational limitations. Indeed, Bucciarelli and Johnson-Laird found that, although their participants drew a slightly more diverse set of conclusions when they constructed external models than when they did not, they were moderately consistent in the conclusions that they drew in the two conditions (60% of their conclusions were logically identical). This result suggests that the performances of the participants, when dealing with the same problems encountered twice, were constrained by nontransitory problems at the algorithmic level. What does make reasoners' performance moderately consistent in the two conditions? And more in general, what does make reasoners differ enormously in their syllogistic ability. In the perspective offered by mental model theory reasoners can perform quite consistently with problems encountered twice because they always rely on their ability to construct and manipulate models of the premises. As construction and search for the integrated models of the premises is supported by the working memory, and individuals differ in their working memory capacity, working memory capacity *can* be considered a principle according to which reasoning diverges from the normative principles. In particular, reasoners with poor working memory capacity are poorer than those with high working memory capacity in solving syllogisms (Bara et al. 1995; 2000). Although variation in working memory is almost entirely captured by measures of general intelligence – as S&W point out – it is predictive of the ability to solve syllogisms. As a consequence, working memory capacity contributes to determine the individual styles of reasoners.

However, in line with an interpretation of syllogistic errors in terms of performance errors, the experimental results of the external models condition showed that the participants differed in which premise they interpreted first, in how they interpreted the premises, and in how they went about searching for counterexamples to the models of the premises constructed initially. In other words, they used different sequences of operations to reach the same result or different results. The most relevant finding, for the present argument, is that these differences occurred not only between individuals, but also within individuals from one problem to another. Certain aspects of the participants' performance were predictable in a probabilistic way, for example the most common interpretation of a premise in a particular mood. But, the results show that it is impossible to predict precisely what an individual will do on a particular trial. This result is consistent with studies in the developmental literature, where it has been found that children use a variety of strategies when reasoning with spatial problems (Ohlsson 1984) and causal problems (Shultz et al. 1986). In particular, children as well as adults tend to use different strategies with the same problem encountered twice.

The methodological implication of my argument is that a theory of reasoning faces the problem of predicting reasoners' performance in spite of the great variability of the strategies they use. Indeed, individuals *seldom* carry out a fixed deterministic strategy in any sort of thinking (Johnson-Laird 1991). As it seems almost impossible to characterize how a certain individual performs in a specific reasoning problem, Johnson-Laird (in Bucciarelli & Johnson-Laird 1999) suggests that an element of nondeterminism must be built into any theory of reasoning. One way to express such a theory would be a grammar with alternative rules that allow for alternative ways in which to represent premises, formulate conclusions, and search for counterexamples. In this way, the theory could be used to "parse" each sequence of models constructed in reaching conclusions. In my view, the introduction of nondeterminism into a theory of reasoning would be the solution to a current debate in the psychology of reasoning, whose focus is the appropriateness of the normative models used to evaluate reasoners' performance. In particular, the expression of a theory of reasoning in the form of a grammar would be the alternative solution to the two extreme positions addressed by S&W, that is, either reading off the normative from the descriptive, or to subtly fine-tune and adjust normative applications based on descriptive facts about reasoning performance.

ACKNOWLEDGMENTS

I thank Bruno Bara, Philip Johnson-Laird, and Vincenzo Lombardo who criticized earlier versions of this comment. This work was supported by the Ministero dell'Università e della Ricerca Scientifica e Tecnologica of Italy (co-financed project 1999, Reasoning processes and mental models: Analyses of spatial inferences).

Reversing figure and ground in the rationality debate: An evolutionary perspective

W. Todd DeKay,[a] Martie G. Haselton,[b] and Lee A. Kirkpatrick[c]

[a]*Department of Psychology, Franklin and Marshall College, Lancaster PA 17604-3003;* [b]*Department of Psychology, University of Texas at Austin, Austin, TX 78712;* [c]*Department of Psychology, College of William and Mary, Williamsburg, VA 23188.* **t_dekay@acad.fandm.edu haselton@psy.utexas.edu lakirk@wm.edu**

Abstract: A broad evolutionary perspective is essential to fully reverse figure and ground in the rationality debate. Humans' evolved psychological architecture was designed to produce inferences that were adaptive, not normatively logical. This perspective points to several predictable sources of errors in modern laboratory reasoning tasks, including inherent, systematic biases in information-processing systems explained by Error Management Theory.

Stanovich & West (S&W) suggest that one of their purposes is to reverse the figure and ground in the rationality debate. We contend that they have not gone far enough. In this commentary, we extend their main point by offering a broader evolutionary perspective.

Life has existed on this planet for a few billion years, modern Homo sapiens for perhaps 500 centuries, Wason tasks and bank-teller problems for only a few decades. All species that have ever existed have done so by virtue of evolved architectures designed by natural selection to solve survival and reproductive problems in ways that enhanced inclusive fitness relative to available alternative designs. The small subset of species possessing brains, including humans, is characterized by a diverse array of specialized evolved psychological mechanisms. Today we are capable of constructing, and sometimes even correctly solving, novel and clever logical and mathematical problems.

The fact that people often make errors in solving Wason-task and bank-teller problems is not surprising. The astonishing fact is that we can conceptualize such problems at all – much less solve them. Natural selection favors organic designs that outreproduce alternative designs, not necessarily those that are normatively rational. In the currency of reproductive success, a rational inference that ultimately interferes with survival or reproduction always loses to a normatively "flawed" inference that does not. This means that good decision-making will not necessarily conform to the rules of formal logic.

"Panglossian" attempts to find explanations for reasoning errors that "prevent the ascription of irrationality to subjects" (S&W, sect. 1, para. 2) have clearly missed this point. The "Meliorist" claim that "human cognition [is] characterized by systematic irrationalities" – as if humans are fundamentally irrational and only sometimes are accidentally rational – is equally misdirected. We contend that it is most useful first to try to identify our evolved psychological mechanisms and their proper functions, and then attempt to determine how humans manage to recruit these mechanisms for purposes other than those for which they were designed. An evolutionary understanding of the mechanisms used in a given task will provide principled criteria with which we can determine what constitutes a functional error in reasoning (maladaptive decisions), as opposed to adaptive deviations from normative rationality.

From this perspective, there are several reasons why we might expect evolved minds to make errors in laboratory reasoning problems. First, people bring into the lab a toolbox packed with specialized implements, none (usually) designed expressly for the task at hand. The surprise occurs when a task *does* happen to fit a tool with which we come equipped, as when a Wason task problem is couched in social-contract terms (Cosmides 1989).

Second, logic problems typically require one to ignore vast amounts of real-world knowledge to focus entirely on abstractions. The gambler's fallacy, for example, requires the subject to put aside knowledge that many events in the world naturally occur in cyclical causal patterns (e.g., weather patterns, resource distributions), so that a long period of not-X really is predictive that X is overdue (Pinker 1997).

Third, as noted by S&W, our evolved psychological mechanisms succeeded or failed historically not as a function of their adherence to symbolic logic, but rather as a function of their effects on survival and reproduction. Rabbits freeze or flee immediately in response to any sudden sound, apparently mistaking benign events for predators. These systematic errors result from an adaptive system. Ancestral rabbits that were not paranoid did not become rabbit ancestors.

According to Error Management Theory (EMT; Haselton et al. 1998), when the costs and benefits of false negative and false positive errors differed recurrently over evolutionary history, selection favored psychological mechanisms biased toward making the less costly (or more beneficial) error. Optimal designs are not necessarily the most accurate. Some "errors" exist because they helped humans reason effectively within ancestral environmental constraints.

EMT explains some biases and errors in human inference that might otherwise be wrongly attributed to computational limitations or design flaws. Men, for example, tend to overestimate women's sexual intent when confronted with ambiguous cues such as a smile; women, on the other hand, underestimate men's commitment intent (Haselton & Buss, in press). According to EMT, these systematic errors evolved because men who erred on the side of false positive inferences missed fewer sexual opportunities than men who did not. Women who erred on the side of false negative inferences about men's commitment intent were better able to avoid deceptive men oriented towards uncommitted sex.

In the currency of reproductive success, these inferential biases are smart because they resulted in lower fitness costs, not because they are accurate, rational, or formally logical. Like paranoid rabbits, humans may possess adaptively biased inferential mechanisms. These adaptive inferences, however, count as irrational within traditional perspectives on reasoning. To truly reverse figure and ground in the rationality debate, we must first clarify what constitutes a "good" inference. An evolutionary perspective is essential for this task.

The rationality debate from the perspective of cognitive-experiential self-theory

Seymour Epstein
Psychology Department, University of Massachusetts, Amherst, MA 01003
sepstein@psych.umass.edu

Abstract: A problem with Stanovich & West's inference that there a nonintellectual processing system independent from an intellectual one from data in which they partialled out global intelligence is that they may have controlled for the wrong kind of intellectual intelligence. Research on cognitive-experiential self-theory over the past two decades provides much stronger support for two independent processing systems.

More than two decades ago, I proposed a dual-process theory of personality (Epstein 1973) subsequently labeled "cognitive-experiential self-theory" (CEST). Since then, I have further developed the theory, and my associates and I have investigated it with an extensive research program. The most fundamental assumption in the theory is that there are two modes of information-processing, experiential and rational. The operation of the experiential system is preconscious, automatic, effortless, rapid, minimally demanding of cognitive capacity, intimately associated with affect, holistic, associative, and imagistic, and its outcome is experienced passively (we are seized by our emotions) and as self-evidently valid (experiencing is believing). The operation of the rational system is conscious, verbal, effortful, demanding of cognitive resources, affect free, and relatively slow. It is experienced as volitonal and as requiring logic and evidence to support beliefs. As the experiential system in humans is essentially the same system used by higher-order nonhuman animals to adapt to their environments by learning from experience, it has a very long evolutionary history. In contrast, the rational system is a verbal inferential system with a very brief evolutionary history.

The two systems operate in parallel and are interactive. All behavior is assumed to be influenced by a combination of both systems, with their relative contribution varying from minimal to maximal along a dimension. The systems usually interact in such a harmonious, seamless manner that people believe they are operating as a single system. The combined operation of the two systems usually results in compromises, but sometimes it produces what are commonly identified as conflicts between the heart and the head.

A paradox with implications for the continued existence of the human species is that human thinking is often highly irrational, to the point of being seriously destructive to self and others, despite the human capacity for very high levels of rational thinking. Why is it that people can solve the most complex technological problems, yet often fail to solve much simpler problems in living that are more important to their existence and personal happiness? The answer, according to CEST, is that the operation of the rational mind is biased by the operation of the experiential mind. It follows that the only way people can be truly rational is to be in touch with their experiential mind and take its influence into account. This is not meant to suggest that the rational mind is always superior. The inferential, rational mind and the learning, experiential mind each has its advantages and disadvantages. Sometimes the promptings of the experiential mind, based on generalizations from past experience, are more adaptive than the logical reasoning of the rational mind.

Our research program has provided support for almost all of the above assumptions. We devoted particular attention to the assumption that there are two independent information-processing systems. Accordingly, I read the target article by Stanovich & West (S&W) with considerable anticipation, hoping to find new, important information in support of dual-process theories. Unfortunately, I found the evidence they cited disappointing.

S&W cite two kinds of evidence from their own research in support of dual-process theories. The first is that, with variance of intellectual intelligence controlled, there are significant positive intercorrelations among responses to a variety of problems in the heuristics and biases literature. S&W argue that this common variance indicates the existence of a broad, nonintellectual cognitive ability. The difficulty with their argument is that it does not take into account that intellectual intelligence is not all of a piece. Thus, it is possible that, had they partialled out the appropriate group factor of intellectual ability, there would not have been any significant systematic variance left. I am not saying I believe that this is what would actually happen, only that it is a possibility that needs to be ruled out before their conclusion can be accepted.

The other evidence S&W cite consists of relations between questionnaire responses and performance on problems from the heuristics literature, controlling for intellectual intelligence. This relation is here stated without elaboration. The same problem concerning group intelligence factors holds as for the other research. Moreover, without information on the nature of the questionnaire responses or a discussion of the meaningfulness of the relations from a coherent theoretical perspective, it is impossible to evaluate their broad statement. They may be able to supply such information, but they have not done so in the present target article, perhaps because of space limitations.

The research my associates and I have conducted provides much stronger support for two processing systems. In the development of a self-report questionnaire on individual differences in thinking style, we found that experiential and rational thinking are not opposite ends of a single dimension. Rather, they are uncorrelated with each other, and they establish different coherent patterns of relations with other variables. We obtained similar findings between a measure of the intelligence of the experiential system and a measure of intellectual intelligence. The measures were unrelated to each other and produced different and coherent patterns of relations with a variety of variables. In studies with problems from the heuristics literature, we found that people who gave heuristic responses were often able to give rational responses when instructed to do so. With the use of a unique experimental paradigm that presents a conflict between the two processing modes in the context of a game of chance in which significant amounts of money are at stake, we found that most participants responded nonoptimally and according to the principles of the experiential system, while acknowledging that they knew it was foolish to bet against the probabilities. Yet they only behaved irrationally to a modest extent, producing responses indicative of compromises between the two processing modes. Space limitations do not allow me to present more extensive and detailed information on the research program. Further information can be found in a number of published articles (for recent reviews, see Epstein 1994; Epstein & Pacini 1999).

Fleshing out a dual-system solution

James Friedrich
Department of Psychology, Willamette University, Salem, OR 97301
jfriedri@willamette.edu

Abstract: A prospective integration of evolutionary and other approaches to understanding rationality, as well as incorporation of individual difference concerns into the research agenda, are major contributions of Stanovich & West's analysis. This commentary focuses on issues of concern in detailing a dual-system or dual-process model of the sort they propose and using it as a basis for improving judgment.

Stanovich & West's (S&W's) proposed resolution of the rationality issue in terms of a dual process approach – System 1 and System

2 – holds great promise, particularly as a bridge between evolutionary and traditional "bias" perspectives. In many ways, the bridge between evolutionary rationality and individual or instrumental rationality, to use their terms, is similar to the bridge between basic cognitive/sensory research and human factors engineering. In designing an airplane cockpit, one can marvel at the sophistication and phenomenal performance capabilities of the human cognitive and perceptual system, but one must also confront the fact that its "normal" performance can be catastrophic in certain circumstances. A vigorous defense of the general utility of humans' basic construal and performance strategies must avoid the excess of implying that any resulting catastrophe is of no consequence – that all "crashes" are somehow unimportant or deniable.

I suspect, however, that the interaction of the two systems is likely to be far more subtle and complicated than is typically implied by most dichotomies. In particular, the notion that System 2 might essentially function as an "override" for System 1 poses a number of research challenges. First, it seems likely that the role of System 2 activity is systematically overestimated in the research literature. For example, as S&W note, cognitive ability predicts performance primarily when System 1 and System 2 processes cue different responses. But it is unclear how common such differential cuing is outside of controlled laboratory conditions. And even for carefully devised experimental tasks, one cannot readily infer System 2 processes on the basis of responses that conform to normative standards. In many cases, successful performance may simply be a byproduct of System 1 processes (cf. Friedrich 1993).

The pervasiveness of System 2 processing may also be overestimated by virtue of the populations studied. To the extent that a disproportionate amount of the published literature emanates from elite institutions studying their own student populations, "g's" positive manifold with performance on many bias tasks suggests that performance base rates, and the manner in which System 2 is typically deployed, could be easily mischaracterized. Combined with situational demands for "high g" research participants to be on their best behavior and to provide the kind of analytical responses they might think a scientist/researcher values, it is easy to obtain distorted pictures of typical performance and System 2 engagement.

A second critical element of a dual process account concerns the manner in which System 2 interacts with System 1. The notion of an "override" system would suggest a fairly clear independence of the two. But the automatic, non-conscious nature of System 1 likely contaminates System 2 (cf. Wilson & Brekke 1994). Conscious, deliberative processes may act primarily on the outputs of more primitive System 1 processes. Thus, certain cognitive structures, values, and preferences on which controlled processing might be performed are themselves often comprehended vaguely if at all at any conscious level (Bargh & Chartrand 1999). One consequence of this is that System 2 overrides may lead to sub-optimal decisions in certain circumstances. For example, Wilson and colleagues (e.g., Wilson et al. 1989) have shown that when people are asked to give reasons for their attitudes, the subsequently generated attitudes seem to be less reflective of people's evaluative responses and dispositions to act. Conscious, deliberative processes may seek to explain or justify preferences that are dimly understood. In this process of "sense making," they will not necessarily do this accurately or in ways that serve the individual's goals.

Perhaps one of the most far-reaching contributions of the paper is the reintroduction of individual differences into the picture. Although such differences have typically been treated as error variance in cognitive research, S&W's use of the positive manifold establishes their significance in answering the question of normativeness and, at least by implication, in addressing concerns of melioration. Of perhaps equal or greater importance than cognitive ability for understanding the operation of a dual system model, however, are motivational differences explored more thoroughly elsewhere by these and other authors (e.g., Cacioppo et al. 1996; Stanovich 1999). In essence, motivational differences are central to the question of when System 2 capabilities will be deployed.

Although motivationally relevant "thinking dispositions" have been linked to the engagement of System 2 processes (e.g., Sà et al. 1999; Stanovich & West 1997), a critical element only indirectly addressed in this literature is the perceived *instrumentality* of analytical processing. Motivation to engage System 2 capabilities is more than just a preference for thoughtful reflection, as reflection and contemplation can occur in either highly contextualized or decontextualized ways. Specifically, it may be the case that even among individuals high in System 2 capabilities, some will not perceive analytical, decontextualized processing as instrumental in bringing about better outcomes. Perhaps the prototypical case example is the psychology major receiving methodological training. A student may dutifully learn and master principles of good design and scientific inference but remain unconvinced that such approaches are truly instrumental in achieving "better" understandings of behavior. Thus, apparent failures to generalize such training beyond an "exam" may reflect an epistemological stance of sorts; if System 2 capabilities exist but are not viewed as helpful, such capabilities are unlikely to be invoked (cf. Friedrich 1987).

In terms of melioration, then, reducing cognitive bias and enhancing critical thinking may require more than adapting the environment to human capabilities, augmenting those capabilities, and enhancing people's pleasure in contemplative, open-minded thought. Persuading people of the instrumental value of decontextualized System 2 processing would also seem critical, though perhaps among the hardest of things to accomplish. For example, the mere presence of simple and easily processed quantitative information in arguments has been shown to shift people from analytical to more heuristic-based processing (Yalch & Elmore-Youch 1984). Highly contextualized System 1 processes are generally quite effective for dealing with the vicissitudes of daily life. Add to this certain self-serving attributions regarding one's naive judgment strategies and an imperfect feedback system that cancels and reinforces many reasoning errors, and instrumentalities for System 2 processing might be very difficult to strengthen. Nevertheless, a thorough treatment of such motivational elements is essential for deriving the full conceptual and corrective benefits of a dual-system understanding of rationality.

The tao of thinking

Deborah Frisch
Department of Psychology, University of Oregon Eugene, OR 97403-1227
dfrisch@oregon.uoregon.edu

Abstract: I discuss several problems with Stanovich & West's research and suggest an alternative way to frame the rationality debate. The debate is about whether analytic (System 2) thinking is superior to intuitive (System 1) thinking or whether the two modes are complementary. I suggest that the System 1/System 2 distinction is equivalent to the Chinese concepts of yin and yang and that the two modes are complementary.

A common finding in research on judgment and decision making (JDM) is that people's responses to simple reasoning tasks often conflict with the experimenters' definitions of the correct answers. Researchers disagree about whether this discrepancy is due to a flaw in the subjects or to a flaw in the experimenters' standards. Stanovich & West (S&W) have demonstrated that subjects who agree with the experimenters on the most controversial questions tend to have higher SAT scores than those who do not. They argue that this supports the claim that the flaw is in the subjects. I shall discuss why I disagree with this conclusion.

First, if S&W's theory is correct, they should expect to find the same correlation in a replication of their studies with JDM researchers as subjects. That is, their analysis predicts that JDM researchers who challenge the conventional wisdom are less intelligent than researchers who accept the standard answers to reasoning tasks. However, it is extremely unlikely that the most vocal critics of the conventional wisdom – Allais (1953), Ellsberg (1961), Gigerenzer (1991b; 1996a) and Lopes (1981b; 1996) are less intelligent than other researchers in the field.

S&W's claim is also weakened by the fact that SAT score only predicts performance on "controversial" problems. For example, on the controversial "destination" version of the four card problem, the effect of cognitive ability is "diluted" (sect. 6). There are two problems with this explanation. First, if both modes cue the same response on the uncontroversial problems, you would expect more consensus on these problems than on the controversial ones. However, there is less consensus on the uncontroversal problems than on the controversial ones. Second, S&W claim that the source of variability on controversial problems is intelligence and the source of variability on uncontroversial problems is some other unspecified factor. However, they do not give a reason why there are two sources of individual differences on these problems. Thus, their account is incomplete and not parsimonious.

S&W describe the two camps in the rationality debate as "Panglossian" versus "Meliorist." This frame distorts the "non-Meliorist" perspective. Specifically, they assume that the main reason researchers "reject the norm" is because many subjects do. However, several researchers have rejected the conventional norms for intellectual reasons. For example, Allais (1953) and Lopes (1981b; 1996) have argued that it is perfectly rational to violate the independence axiom of utility theory. Similarly, Kahneman and Tversky (1984) and Frisch and Jones (1993) have argued that violations of description invariance (framing effects) can be sensible.

The target article suggests an alternative way to frame the rationality debate. In Table 3, S&W describe a distinction between an intuitive mode of reasoning and an analytic one. The debate among JDM researchers boils down to the question of how these two modes are related. The "Meliorist" view assumes that the analytic mode is superior to the intuitive one. The "non-Meliorist" view does not assume that either mode is superior to the other.

A better term for the "non-Meliorist" view is the "complementary view." I suggest this term because of a striking similarity between S&W's System 1/System 2 distinction and the Chinese distinction between yin and yang. Capra (1982) describes yin thinking as intuitive, synthetic, and feminine and yang thinking as rational, analytic, and masculine. This is essentially the same as the System 1/System 2 distinction. In S&W's view, yang is superior to yin. But in Chinese thought, the two modes are complementary. As Capra (1982) says, "What is good is not yin or yang but the dynamic balance between the two; what is bad or harmful is imbalance" (p. 36). A growing number of JDM researchers including Epstein (1994), Hammond (1996), and Klein (1998) have endorsed views similar to the complementary view.

In the "Meliorist" versus "Panglossian" frame, only the Meliorists offer advice for improving the quality of thinking. The Panglossians think things are fine the way they are. In the Meliorist (or classical) versus complementary frame, both sides acknowledge it is possible to improve the quality of thinking but they offer different advice about how to achieve this goal. Advocates of the classical view believe that to improve the quality of thinking, a person should increase the extent to which he relies on the analytic mode. Advocates of the complementary view believe that improving the quality of thinking involves achieving an integration between intuitive and analytic processing. In fact, on the complementary view, becoming more analytic can be detrimental if a person was already out of balance in that direction to begin with.

S&W endorse the classical view, while I favor the complementary view (although I realize it is not really in the spirit of complementarity to pit the two views against each other). In closing, I would like to comment on a novel justification provided by S&W for the superiority of System 2 (yang) over System 1 (yin). In section 6.3, they say that "'Life,' in fact, is becoming more like the tests!" (p. 35). The idea is that in our increasingly technological society, the analytic mode is more adaptive than the intuitive mode. However, while it is true that the human made world is increasingly dominated by the abstract, analytic mode (e.g., computer technology), the natural world is *not* becoming more like the tests. To the extent that S&W really believe that "life" is equivalent to "the human made world," they have provided an excellent example of the dangers of excessive reliance on yang thinking.

Gone with the wind: Individual differences in heuristics and biases undermine the implication of systematic irrationality

David C. Funder

Department of Psychology, University of California, Riverside, Riverside, CA 92521 **funder@citrus.ucr.edu**
www.psych.ucr.edu/faculty/funder/RAP/RAP.htm

Abstract: The target article's finding of stable and general individual differences in solving of problems in heuristics-and-biases experiments is fundamentally subversive to the Meliorist research program's attention-getting claim that human thought is "systematically irrational." Since some people get these problems right, studies of heuristics and biases may reduce to repeated demonstrations that difficult questions are hard to solve.

The target article's analyses of individual differences in susceptibility to cognitive errors pose two serious difficulties for the "Meliorist" program of research on human judgment, even though Stanovich & West (S&W) seem at pains to deny them. The first difficulty is relatively minor; it applies only to some Meliorist studies (though more than one might have expected). The second undermines the foundation of the whole program or, if not its foundation, then its most attention-getting aspect.

First, the minor difficulty. Individual difference data show that people who give the "wrong" answers to Meliorist problems are not always less smart or less successful than those who get them right, and therefore may not be manifesting flawed reasoning. In my own research, subjects who committed a well-studied attributional error appeared *socially advantaged* in nonlaboratory contexts, suggesting that the error manifested a social competence rather than a deficiency (Block & Funder 1986). Findings like these are reminiscent of a midterm exam where a professor discovers that a question either has a flat response curve (reflecting that better students, on the other questions, were no more likely to get this one right), or an inverted one (indicating that better students were more likely to get this item wrong). In the first circumstance, the probable implication is that the professor has written a poor item that fails to discriminate the better and worse students. In the second circumstance, the implication is that the wrong answer was keyed as correct. So it is when one finds, in a Meliorist study, that smarter subjects either do not provide the normative answer or even provide a different answer: the question should be tossed out or re-keyed.

This keying problem indeed seems to afflict a few of the more prominent errors discovered by Meliorist research, such as the putative underutilization of noncausal base rates. But for purposes of further discussion, we can simply set such phenomena aside. With a few notable exceptions, Meliorist studies are probably keyed correctly, the smarter and more competent people are more likely to get the advertised correct answer. However, this finding raises a second problem that is even more ominous for the Meliorist position.

According to the target article (sect. 1, para. 2), the key claim of Meliorist research is that "human cognition [is] characterized by systematic irrationalities."[1] This attention-getting claim is precisely what led Meliorism to become rich and famous. Its very power and appeal stems from the idea that human thought is *sys-*

tematically irrational; a fundamental shortcoming of the architecture of the human cognitive system causes its inferential processes inevitably to go awry. If this idea is true, the implications are profound.

But the finding of broad and stable individual differences in the susceptibility to bias is powerfully subversive to this idea. It implies that the (correctly keyed) errors assessed by Meliorist research reveal not systematic irrationality, but variations in the ability to answer difficult questions. Some questions are so difficult that only very smart people get them right. The Wason task, for example, is hard to figure out (depending, actually, on how it is worded). A few manage to solve it; most do not. And when you give people on the street certain problems in syllogistic reasoning, statistical inference, or covariation detection, most will probably get them wrong but again, a few smart ones will surprise you.

The presence of people – even a few people – who consistently do not miss Meliorist problems implies that what errors demonstrate is not some fundamental limitation on human rationality, but something akin to what the Educational Testing Service (ETS) demonstrates every time it writes a difficult SAT item. As far as I know, nobody has ever claimed that the existence of SAT items that most test-takers get wrong means that human cognition is systematically irrational. Yet this is precisely the kind of implication drawn by Meliorism every time it interprets the invention of a difficult problem as revealing a fundamental limitation on human thought. The target article thus – intentionally or not – exposes the untenability of Meliorism's most dramatic and attention-getting claim. I believe this is the reason why Melioristically inclined reviewers of earlier drafts were so upset with it, despite the authors' yeoman attempts (increased in this final version) to deny any such subversive intent.

Near the end of the target article, S&W note that "significant covariance among the scores from the variety of tasks in the heuristics and biases literature [remains] after they [have] been residualized on measures of cognitive ability" (sect. 7, para. 3). In other words, Meliorist problem-solving is determined by more than just IQ. If this observation is intended to rescue the implication of systematic irrationality, it falls far short, for three reasons. First, the target article's Table 1 demonstrates the impressive amount of variance that individual differences in Meliorist reasoning tasks do share with (of all things) SAT scores – so much so that many Meliorist reasoning tasks would make pretty decent SAT items and perhaps should be brought to the attention of ETS! Second, "residualizing" measures of cognitive ability is limited in its effect to the reliability (always less than perfect) of the measures used and, moreover, removes the influence only of the specific (and sometimes narrow) cognitive skills that they happen to tap. Intelligence, as has been widely publicized recently, is much more than what is measured by typical tests of cognitive ability.

Third, and most important, the existence of significant, stable, and general individual differences in problem-solving on Meliorist tasks – regardless of what turns out to correlate with those differences – implies that the vast literature on heuristics and biases may embody little more than a collection of brain teasers that most people get wrong but that a few people – without tutoring and despite everything – manage to get right. The Meliorist research program might still be worth pursuing, to the extent that it can show how improvement at the kind of problem-solving it assesses has benefits for reasoning or daily living. And a close analysis of the kinds of errors people tend to make, when they make them, might be helpful in the design of decision-making aids. But these relatively mundane possibilities are not what made the heuristics and biases approach so famous in the first place. The attention-grabbing notion that the human mind is afflicted by "systematic irrationality" was fun while it lasted, but is gone with the wind.

ACKNOWLEDGMENT
The author's research is supported by National Institute of Mental Health grant R01-MH42427.

NOTE

1. In a phrase that I believe is intended to mean more or less the same thing, the target article states near its end (sect. 7, para. 3) that human reasoning reflects "a systematically suboptimal intentional-level psychology."

Patterns of individual differences and rational choice

Vittorio Girotto
CREPCO, Université de Provence, Aix-en-Provence, 13100, France; and Department of Psychology, Trieste University, Trieste, Italy.
girotto@up.univ-mrs.fr girotto@mail.univ.trieste.it

Abstract: I discuss an aspect of individual differences which has not been considered adequately in the target article, despite its potential role in the rationality debate. Besides having different intellectual abilities, different individuals may produce different *erroneous* responses to the same problem. In deductive reasoning, different response patterns contradict deterministic views of deductive inferences. In decision-making, variations in nonoptimal choice may explain successful collective actions.

I am sympathetic to Stanovich & West's (S&W's) proposal that individual differences may serve as a tool for analysing the gap between normative models and actual performance. S&W have provided convincing evidence that not all errors in thinking problems are owing to what Panglossian authors may label as "adventitious causes." There is, however, an aspect of individual differences that S&W have not considered adequately, despite its potential role in the rationality debate.

In the thinking and reasoning literature, individual differences appear to coincide with differences in general intellectual abilities. S&W have extensively investigated this aspect of individual differences. Showing that more intelligent individuals perform better than less intelligent ones in some reasoning problems is clearly relevant for evaluating the rationality of human thinking. Individuals, however, vary: not only in their general cognitive abilities. As S&W note in their concluding remarks, individuals also differ in their thinking styles (e.g., they may have different dispositions toward confirmation, premature closure, etc.). Following mental model theory, S&W hypothesise that these factors may determine the extent to which reasoners search for potential counterexamples, that is for contradictory models of the premises. Indeed, some empirical results corroborate the non-deterministic stance of model theory (i.e., the assumption that a given set of inputs may elicit different representations and different responses by different individuals). Reasoners appear to produce different answers (all equally *incorrect*) to the same reasoning problem. For example, some probabilistic reasoning problems elicit different inferences which may be owing to different representations of the same premises (e.g., Girotto & Gonzalez 2000; see also Stanovich & West 1998c). These and other results obtained with deductive (e.g., Bucciarelli & Johnson-Laird 1999) and meta-deductive (e.g., Johnson-Laird & Byrne 1990; 1993) reasoning tasks contravene the deterministic assumption according to which individuals use just one general strategy to solve various reasoning problems (see Rips 1989).

Variations in the patterns of nonnormative responses are relevant for a further reason. According to rational choice theory, individuals tend to maximise their expected utility. Actual decision patterns, however, may depart from economic rationality. For example, many individuals solve conflicts in decision making by acting cooperatively, in the service of collective interests, rather than competitively, in the service of individual interests. Thus, the extensive literature on one-shot conflicts with the structure of a Prisoner's Dilemma (PD) game show that between one-third and one-half of the participants cooperate (see Sally 1995). How can cooperation in one-shot PD games be explained? We may argue that cooperation is due to some "adventitious causes," that is, individuals may interpret one-shot games as repeated games (in which, under some parameters, a cooperative strategy is rational in the sense that it optimises one's long-run outcome, see Axelrod 1984). In sum, we may apply the alternative construal argument to cooperation in single-round mixed-motives games. There are theoretical (e.g., Elster 1989) as well as empirical (e.g., Shafir & Tversky 1992) reasons to doubt that all cases of cooperation are due to an inappropriate interpretation of the game. In real life, moreover, people do cooperate even when they know that no retaliation would result from competitive behaviour (e.g., Mansbridge 1990).

Contrary to the assumption that there is a norm of cooperation, which should tautologically explain the existence of cooperative choices, empirical results support Elster's (1989) hypothesis that successful collective action is produced by a mix of individual tendencies. In particular, Morris et al. (1998) showed that cooperation in one-shot PD games is due to various sources. On the one hand, some individuals follow a decision heuristic to "match" the counterpart's possible cooperation. Response patterns due to this heuristic, which is related to the social norm of reciprocity, are more frequently elicited when the counterpart's move lies in the past than when it lies in the future. On the other hand, other individuals follow a "control" heuristic, that is they cooperate *as if* their move could determine the counterpart's cooperation. Response patterns due to this heuristic, which is related to the cognitive difficulty of reasoning under uncertainty (Shafir & Tversky 1992), are more frequently elicited when the counterpart's move lies in the future than when it lies in the past. These different patterns of nonoptimal choice may lead to collectively optimal decisions in many interpersonal conflicts. In particular, in games in which the actions of the players cannot be simultaneous, individuals who cooperate on the basis of the control heuristic may act earlier and catalyse the counterpart's cooperation. Conversely, individuals who follow the matching heuristic will be forced to reciprocate the possible "good faith" cooperation of the earlier players. Such an interaction of individual tendencies may yield successful collective actions. If only one of the heuristics existed, nonsimultaneous mixed-motive conflicts would not be resolved efficiently. From a theoretical point of view, these results imply that it will be impossible to construct a general theory of collective action "The variety of interacting motivations is simply too large for any equilibrium theorem to be proved" (Elster 1989, p. 205).

In summary, just as variations in erroneous patterns in reasoning and judgement problems are relevant for testing Panglossian assumptions about human thinking, variations in nonoptimal choice patterns may be relevant for testing rational choice assumptions about human decision-making.

Some theoretical and practical implications of defining aptitude and reasoning in terms of each other

Adam S. Goodie and Cristina C. Williams

Department of Psychology, University of Georgia, Athens, GA 30602-3013
{goodie; williams}@egon.psy.uga.edu
http://teach.psy.uga.edu/Dept/Faculty/Goodie.stm

Abstract: Stanovich & West continue a history of norm-setting that began with deference to reasonable people's opinions, followed by adherence to probability theorems. They return to deference to reasonable people, with aptitude test performance substituting for reasonableness. This allows them to select independently among competing theories, but defines reasoning circularly in terms of aptitude, while aptitude is measured using reasoning.

Stanovich & West (S&W) propose that aptitude tests may be the best arbitrators of normative models. Applying it judiciously, they bring a welcome measure of reasonableness (one might even say rationality) to the rationality debate. However, aptitude tests and questions that involve reasoning ability are not independent: the latter define a large part of the former, so certain circularities must be addressed when the former are proposed as a definition of the latter.

S&W's central question is: Who or what is the final authority on correct solutions to problems of uncertainty. As Gigerenzer (1991b) has pointed out, the classical probabilists relied on the common wisdom of reasonable people. When the judgments of reasonable people differed from the conclusions of probability theory, it was the reasonable people's judgment that prevailed, and it was the probability theory that was adjusted to resolve the conflict. The qualifier "reasonable" in this context serves to prevent a few nonconventional thinkers from undoing a consensus. It could conceivably have altered the development of probability theory – for example, an elite group of probabilists could proclaim themselves the only reasonable thinkers – but nothing of this sort seems to have happened. In general, it was a rather democratic way of deciding the normative status of possible answers to difficult questions.

Later, the probability theories themselves became the normative standard, displacing the judgment of reasonable people (Gigerenzer 1991b). It was the mathematical formulas devised by the probability theorists, and not the theorists themselves, that dictated normative processes. The new norm is less democratic than the previous one, depending as it does on the derivations of a few theorists rather than the collective wisdom of all reasonable people. But it replaced democracy with objectivity, promising that assessments of rationality would be based on universal laws rather than individual opinion.

Unfortunately for this program, different probabilistic theories, or different applications of a single theory to a situation, may produce different norms, and there is no universal law to dictate which one is right. These complications have helped to foster a backlash against probability theory (e.g., Gigerenzer & Todd 1999, this issue) and a renewed dependence on people's opinions, but it has become clear that those with an interest in these issues seldom share a single, "reasonable" opinion about any of them. Whom shall we believe?

S&W aptly note that debates concerning which normative model to adopt are linked to the observed distribution of models used by the untutored population. New models are not promoted on purely rational bases, but rather with an eye toward what they say about human abilities. However, as S&W also note, re-construing the task or suggesting an alternative normative model for an accepted task is done only by those who defend human rationality, when they encounter a case in which people seem at first glance to behave sub-optimally. The same sort of thing could be (but is not) done by skeptics of human rationality to reinterpret an initial finding of rational or adaptive behavior. S&W protest too much, however, against the elitist/populist distinction. Because some people provide the correct answer on these tasks, it is true that the difference is between theorists and some laymen on the one hand, and other laymen on the other hand; it is not simply between theorists and laymen. But it must be emphasized that in the empirical domains where these debates erupt, the Meliorists stand in opposition to *most* laymen. If it were otherwise, then there would be no need for the Panglossians defending human rationality against the Meliorists; they could simply point to the rationality of the average layman as the rational male exemplar.

S&W are taking a bold step in addressing the prescriptive issue of adjudicating among normative models. They note that aptitude test scores correlate with different application strategies, and in particular that the strategies of the Meliorists tend to be selected by people who achieve high scores on aptitude tests (henceforward, "smart" people, although their sampling of aptitude tests in the target article is restricted to the SAT). This is taken as justifying the Meliorists' applications, and supporting the Meliorists' view that people suffer from systematic biases. Consistently, in the few instances where aptitude test scores correlate negatively with a Meliorist view of a particular problem, notably in the domain of noncausal base rates (e.g., Goodie & Fantino 1999), the Panglossian view is adopted.

Normative status is thus conferred to the answers provided by smart but untutored people. This is a significant and creative step in the development of normative authority, but it involves a certain circularity. Aptitude test scores are, after all, based largely on the frequency with which one provides normatively correct answers to reasoning problems. Thus, smart people are those who provide normative answers, and normative answers are those provided by smart people. They are defined in terms of each other precisely because of the decision to credit the answers of smart people as being definitively normative. This is a curious outcome indeed, and one with a potential for ill effects. A "Tyranny of the Bayesians" could emerge, or a "Signal Detection Regime," that perpetuated itself by claiming that it was not they but the smart laymen who decided what means should be used to make decisions under uncertainty, while they surreptitiously manipulated which untutored people would appear smart on tests of aptitude. Of course, there is already widespread opinion to this effect regarding the determination of aptitude tests (e.g., Sternberg & Wagner 1993). Stanovich & West's proposal is a bold new step with significant implications for both theories of reasoning and the role of aptitude tests. They should address the ostensible circularity of creating modes of rationality and aptitude tests that depend heavily on each other.

Individual differences: Variation by design

Anthony J. Greene & William B. Levy

Department of Neurosurgery, University of Virginia, Charlottesville, VA 22908
ajg3x@virginia.edu www.people.virginia.edu/~ajg3x

Abstract: Stanovich & West (S&W) appear to overlook the adaptivity of variation. Behavioral variability, both between and within individuals, is an absolute necessity for phylogenetic and ontological adaptation. As with all heritable characteristics, inter-individual behavioral variation is the foundation for natural selection. Similarly, intra-individual variation allows a broad exploration of potential solutions. Variation increases the likelihood that more optimal behaviors are available for selection. Four examples of the adaptivity of variation are discussed: (a) Genetic variation as it pertains to behavior and natural selection; (b) behavioral and cognitive aspects of mate selection which may facilitate genetic diversity; (c) variation as a strategy for optimizing learning through greater exploration; and (d) behavioral variation coupled with communication as a means to propagate individually discovered behavioral success.

Variation is often misconstrued to be nonadaptive because it insures that many individuals will be less than optimally suited to the environment. For example, Karl Pearson (1897) speculated that cognitive performance should be homogeneous due to a disadvantage to those individuals who deviate substantially from some optimum. While strong selection pressures may exert a proximal constraint on variability, environmental change exerts a universally opposing selection pressure for variation (e.g., Slatkin 1974). "These individual differences are highly important for us, as they afford materials for natural selection to accumulate" (Darwin 1859/1959, p. 123). Indeed, mechanisms for enhancing variation constitute the principle adaptation which made virtually all subsequent adaptations possible: The capacity to produce a unique combination of genes in each individual and at each generation (namely, sexual reproduction) is the point of divergence between the narrow range of cloning species (e.g., amoebae) and all other variety of life. Reproductive behavior itself may have a strong influence on genetic diversity: One line of evidence for this claim is that female learning varies over the ovarian cycle in ways that tend to maximize the genetic diversity of offspring: During proestrus, changes observed in hippocampal connectivity may underlie novel mate selection while the hippocampal state during diestrus may better serve foraging. The combination of a good memory for a former mating partner coupled with a tendency toward promiscuity during proestrus may be one of the most powerful mechanisms for insuring genetic variability (see Desmond & Levy 1997). Another line of evidence supporting this claim is that twin studies show that mate preference has a negligible heritability component (Lykken & Tellegen 1993) especially compared to other preferences (job, religion, political affiliation, etc.). High variability allows species to occupy novel and unpredictable environments and to survive local and global ecological changes (e.g., Cooper & Kaplan 1982). Genetic continuation is clearly more important than the optimal function of any individual within that species (e.g., Yoshimura & Clark 1991). Thus, the phylogenetic adaptiveness of variation completely circumscribes considerations of local optimizations.

The volatile environment that favored phylogenic variation also favored adaptations that allow organisms to learn, and thereby to survive and proliferate under a much broader range of conditions. While inter-individual variation provides the basis for natural selection, intra-individual variation provides a basis for behavior selection (e.g., Skinner 1981) by increasing the likelihood that more possible problem solving strategies are explored. Several examples illustrate the utility of behavioral variation.

The well-documented phenomenon of "probability matching" (Estes & Straugham 1954) is a clear example where response variation is ostensibly sub-optimal, but provides a basis for response selection. That is, if the likelihood of reward is divided unevenly between alternatives, optimal reward would be obtained by always selecting the highest-reward alternative; however, both animals and humans systematically sample the lower-reward alternatives, thus diminishing their expected reward. This is a senseless choice unless it is remembered that information has value as well (Dinsmoor 1985). In natural environments, the probability of finding food in particular patches is not stable, so that consistent responses (optimizing consumption) result in ignorance of changes in relative probability densities and possibly starvation as the existing repertoire of patches is exhausted. More interesting, sated animals tend to engage in more exploration, while deprived animals tend to maximize reward (McDowell & Dallery 1999). The exploratory value of variation is further supported by experiments showing that the extent of initial variation predicts the likelihood that the problem is later solved. For example, in a number conservation task, five-year-old children who engaged in a wider range of initial strategies were more likely to later converge on the correct solution (Siegler 1995). Similarly, initially high variations in neural network activity (cf. simulated annealing) improve performance by increasing the portion of phase space that is sampled (e.g., Shoal & Hasselmo 1998). Ontologically, the advantage of wide exploration is greatest for juveniles because fewer possibilities have previously been explored. Indeed, Wenger and McKinzie (1996) review numerous findings showing that intra-individual behavioral variation is inversely correlated with age and experience.

Among animals capable of learning from one another, variations in individual performance provide a broader palate for behavior selection by observation. For instance, honeybees explore widely and then communicate successes to their hives (e.g., Seeley 1994). Similarly, primates have been widely observed to communicate effective tool use and foraging strategies to juveniles (for a review see Galef 1998). A wide assortment of behaviors with varying degrees of success, coupled with propagation by either observational sharing or communication can improve the success of an entire clan.

Variation is central to any concept of behavioral or phylogenetic optimization. It is then surprising that the target article reviewed evidence that evolutionary considerations may indicate a shift in the construal of optimal performance, but did not account for variation itself as an evolutionary consideration. While we agree with Stanovich & West that individual variation cannot be accommodated within the perspective of perfect rational competence, we disagree that this constitutes a shortcoming. Adaptivity, both within and between lifetimes, requires a degree variability that is not correlated with the adaptive pressures or task demands of the immediate environment.

The understanding/acceptance principle: I understand it, but don't accept it

David Hardman

Department of Psychology, London Guildhall University, Calcutta House, London E1 7NT, United Kingdom **dhardman@lgu.ac.uk**
www/gu.ac.uk/psychology/hardman/hardman.html

Abstract: Can the understanding/acceptance principle help us to decide between alternative normative theories? There is little evidence that this principle can successfully be applied; there are no strong normative statements in Stanovich & West's target article. There is also no evidence for success of rational norms when applied to real life decisions.

> Why, sometimes I've believed as many as six impossible things before breakfast.
> from *Through the Looking Glass,*
> by Lewis Carroll (1872).

What should we take to be normative? Cohen (1981) famously argued that the intuitions of ordinary people were the appropriate basis for determining rationality. Now Stanovich & West (hereafter S&W) reject the intuition-based approach, on the basis that it is the most intelligent and reflective individuals who are likely to be most in accordance with what is normative. This idea stems from the understanding/acceptance principle; the notion that people are more likely to accept an axiom of logic or probability once they have properly understood it. Thus, in situations where normative models are a matter of controversy, these more reflective individuals can help us decide between those norms. I see several difficulties with this approach, which I describe below.

Can the understanding/acceptance principle be sustained? First, the attribution of the understanding/acceptance principle to Slovic and Tversky (1974) is based on a misreading of that paper, which was concerned with a central axiom of Subjective Expected Utility Theory (SEU). Those authors simply set out to test the view of "[many] theorists . . . that if the axioms were presented in a clear and understandable fashion, no reasonable person would wish to violate them" (p. 368). The data did not support this view, and in a hypothetical dialogue representing the views of Allais and Savage, the issue was raised as to whether or not the subjects could properly have understood the axioms (S&W attribute to Slovic & Tversky certain views that were actually merely part of this hypothetical dialogue).

Nevertheless, one might assume the understanding/acceptance principle regardless of whether Slovic and Tversky endorsed it. Other evidence speaks against the understanding/acceptance principle, however. Denes-Raj and Epstein (1994) found that when attempting to blindly select a red marble from a group of marbles, many participants preferred to select from a saucer containing from five to nine red marbles out of 100 overall (depending on condition), rather than from a saucer containing one red marble out of 10 overall, despite the fact that the ration was more favorable in the latter (5–9% versus 10% red marbles). More important, participants often reported that they knew intellectually that the smaller set of marbles was more favorable, but still felt they had a better chance with the larger set. It appears then that people do not necessarily accept a principle that they understand.

Other examples of failure to apply learned principles comes from the sunk cost literature. Training in cost-benefit principles leads to a better appreciation of those principles (Larrick et al. 1990), but does not necessarily prevent people from committing the sunk cost fallacy (Arkes & Blumer 1985; Larrick et al. 1990), particularly outside of the domain of one's training (Tan & Yates 1995). These examples not only indicate that experts retain certain intuitions, but they call into question the understanding/acceptance principle itself. Problem content seems to be a more influential factor than knowledge of cost-benefit principles.

S&W themselves have found that the most intelligent and reflective people do not always appear to follow traditional normative models. For example, on a base-rate task S&W (1999b) found that the presentation of both normative and non-normative arguments led more people to switch to non-normative responses than to normative ones. What is somewhat curious is that given the extensive data that S&W report, they themselves do not follow through on their basic argument by making strong recommendations about the norms we ought to be using. To me it is not clear that the understanding/acceptance principle is a workable proposition.

Do rational norms bring performance benefits in the real world? A second difficulty is that the current approach fails to show that correspondence to traditional norms brings any benefits outside the psychological laboratory. S&W focus exclusively on the use of familiar lab-based pencil and paper tasks. This is entirely understandable, as such tasks are easy to study and enable the use of large samples. Unfortunately, this approach fails to bridge the gap between the lab and the outside world (e.g., Fischhoff 1996).

If it could be shown, for example, that the application of decision-theoretic principles (i.e., seek out all relevant information, structure options and create options where necessary, consider consequences and uncertainties) to real life situations led to appreciable benefits, then it would be sensible to use them. However, even proponents of decision analysis are unable to point to a body of literature showing the effectiveness of such systematic methods (Clemen 1999). In a number of studies involving real decisions, reflective thinking has been shown to be ineffectual at best, and detrimental at worst. One longitudinal study of a real-life decision (high school students choosing a college) found that students who created more complex decision maps were no more successful in their choice of college, and also were no more satisfied with the decision process itself (Galotti 1995). Another study found that college students were less likely to choose highly-rated courses if they had introspected on their reasons, and they were just as likely as a control group to choose poorly-rated courses (Wilson & Schooler 1991).

Klein (1998) draws on his research in naturalistic decision making to suggest that the use of formal decision methods in certain domains may be positively harmful. He argues that expert decision making largely involves recognizing a situation as the same or similar to previously-encountered situations. Such recognition is likely to trigger the recall of the previous solution. If this solution was satisfactory, then it can be used again without the consideration of other alternatives. Under situations of stress and time pressure, such "recognition-primed" decision making is of utmost importance. This conception of decision making implies that the training of experts should involve the presentation of many decision "scenarios," in order that the expert can learn by example. Training by means of formal decision analysis, on the other hand, could impede the natural acquisition of expertise.

Simon (1983) has argued that SEU can be a useful approximation in some circumstances, but that there may be other models of decision making that lead to better outcomes in the real world. However, as is well known, the assumptions of SEU are unrealistic because they require processing capacities that neither people nor the world's largest computers possess. This has led Simon to state that "SEU theory has never been applied, and never can be applied" (Simon 1983, p. 14). To believe that we should obey the various axioms of an impossible theory is to belong to the looking-glass world of Alice (see quotation at the start of this commentary).

Simon's view that there may be better models of decision making than SEU is clearly beginning to bear fruit. For example, Klein's research shows that experts are "satisficing" in the manner proposed by Simon (1956); that is, they are using "bounded rationality." Other recent work by Gigerenzer and Goldstein (1996) shows that "fast and frugal" heuristics, which use just one reason for making a decision, match or outperform more rational models that integrate various items of information (see also Gigerenzer et al. 1999).

Conclusions. Should untutored intuitions or reflective thinking condition our conceptions of what is rational? S&W are clearly in favour of the latter, but in my view this is premature. Surely what is most important is to know what ways of thinking will provide effective solutions to real world problems. It may be that some intuitions do, in fact, reflect the operation of evolved cognitive mechanisms that do just this (e.g., Gigerenzer et al. 1999), and that other useful intuitions develop through learning (e.g., Klein 1998). But it may also be the case that the modern environment presents us with problems that require more reflective modes of thought. However, determining what is the most effective mode of thought in a given situation requires more than scrutinizing the most intelligent individuals on a series of paper and pencil tasks, which may or may not be representative of real world problems. As we have seen, there are at least some problems in the contemporary environment where reflective thinking does not appear to be particularly helpful.

The questionable utility of "cognitive ability" in explaining cognitive illusions

Ralph Hertwig

Max Planck Institute for Human Development, 14195 Berlin, Germany
hertwig@mpib-berlin.mpg.de
www.mpib-berlin.mpg.de/ABC/Staff/Hertwig

Abstract: The notion of "cognitive ability" leads to paradoxical conclusions when invoked to explain Inhelder and Piaget's research on class inclusion reasoning and research on the inclusion rule in the heuristics-and-biases program. The vague distinction between associative and rule-based reasoning overlooks the human capacity for semantic and pragmatic inferences, and consequently, makes intelligent inferences look like reasoning errors.

Why do most undergraduates appear to get the Linda problem wrong? After all, this problem is meant to instantiate the *inclusion rule,* "perhaps the simplest and most fundamental principle of probability. . . . If A includes B then the probability of B cannot exceed the probability of A" (Kahneman & Tversky 1996, p. 585). Stanovich & West (S&W) (1998b, p. 307) argue that although the problem tests reasoning in accord with a simple rule, "correct responding on the Linda problem . . . is associated with higher cognitive ability." The finding that higher SAT scores are correlated with inclusion responses in the Linda problem is a flagship example of their more general claim that there are two reasoning systems, one associative and the other rule-based, and that students with higher cognitive ability are more likely to give rule-based responses. In what follows, I demonstrate that S&W's use of cognitive ability to explain violations of the inclusion rule, when viewed in light of other findings on reasoning about class inclusion, gives rise to paradoxical conclusions.

Is the cognitive ability of eight-year-olds higher than that of undergraduates? In their classic book *The early growth of logic in the child,* Inhelder and Piaget (1964, p. 101) reported an experiment in which they showed five- to ten-year-old children 20 pictures, four representing colored objects and 16 representing flowers. Eight of the 16 flowers were primulas, four yellow and four of other colors. The children were asked a list of questions about class inclusion relations, one of which was: "Are there more flowers or more primulas?" Only 47% of the five- to seven-year-olds gave answers in accord with class inclusion, that is, which reflected an understanding that the flowers were more numerous than the primulas. Among eight-year-olds, however, a majority (82%) gave responses consistent with class inclusion. Inhelder and Piaget (1964) concluded that "this kind of thinking is not peculiar to professional logicians since the children themselves apply it with confidence when they reach the operational level" (p. 117).

A couple of decades later, Tversky and Kahneman (1983) gave undergraduates at universities such as Stanford the description of a person, Linda, and asked them to rank statements about Linda according to their *probability.* Among them were Linda "is a bank teller" (T) and "is a bank teller and is active in the feminist movement" (T&F). Only 11% of the adult participants ranked T as more probable than T&F, although T&F is included in T. Here we encounter the puzzle. The Linda problem is analogous to the flower problem in that both represent an inclusion relation (Reyna 1991, p. 319). Why, then, do children as young as eight (or nine and ten; Inhelder and Piaget were probably too optimistic about the onset of class-inclusion reasoning; Reyna 1991) follow class inclusion, while undergraduates do not? To the extent that "correct" responding in inclusion problems is associated with higher cognitive ability, as S&W's account suggests, we ought to conclude that eight-year olds have higher cognitive ability than Stanford undergraduates. Not according to Piaget's theory of cognitive development, or for that matter, according to probably any other theory of cognitive development, much less according to common sense: concrete-operational children should trail far behind the undergraduates, who have reached the highest state of cognitive ability, the formal-operatonal stage.

Is the cognitive ability of second graders higher than that of sixth graders? Perhaps the cognitive ability explanation would lead to less paradoxical conclusions if applied only to studies using the Linda and similar problems. In a pertinent study, Davidson (1995) gave second, fourth, and sixth graders problems such as the Mrs. Hill problem. Mrs. Hill "is not in the best health and she has to wear glasses to see. Her hair is gray and she has wrinkles. She walks kind of hunched over." Then, the children were asked to judge how likely Mrs. Hill was to have various occupations, such as Mrs. Hill is "an old person who has grandchildren," and "an old person who has grandchildren and is a waitress at a local restaurant." In Davidson's study, second graders gave more class inclusion responses than sixth graders (65% vs. 43%). Why? If "correct" responding in Linda-type problems is in fact associated with higher cognitive ability, then we ought to conclude that second graders have higher cognitive ability than sixth graders. Again, on any account of cognitive development and common sense, this conclusion is implausible. Ironically, Davidson (1995) interpreted the finding as evidence that children with higher cognitive ability (older children) are more likely to use the representativeness heuristic than children with lower cognitive ability (younger children). Yet the representativeness heuristic seems to epitomize what S&W refer to as the associative system, and thus its use should be correlated with lower cognitive ability.

Why do people violate the principle of class inclusion in the Linda problem? Is there a way out of these paradoxes? In my view, notions such as "cognitive ability," which explain everything and nothing, will not be of much help to us in understanding people's reasoning abilities; nor will "dual-process theories of reasoning," unless underlying cognitive processes are clearly specified (for a critique of such theories, see Gigerenzer & Regier 1996). Problems such as Linda, the cab problem, and the standard Wason selection task are inherently ambiguous (e.g., Birnbaum 1983; Hilton 1995). The Linda problem, for instance, is not a mere instantiation of the inclusion rule. It is laden with the ambiguity of natural language. Take the word "probability." In probabilistic reasoning studies, "probability" is typically assumed to be immediately translatable into mathematical probability. From its conception, however, "probability" has had more than one meaning (e.g., Shapiro 1983), and many of its meanings in contemporary natural language have little, if anything, to do with mathematical probability (see Hertwig & Gigerenzer 1999). Faced with multiple possible meanings, participants must infer what experimenters mean when they use the term in problems such as Linda. Not surprisingly, participants usually infer nonmathematical meanings (e.g., possibility, believability, credibility) because the Linda problem is constructed so that the conversational maxim of relevance renders the mathematical interpretation of "probability" implausible (Hertwig & Gigerenzer 1999). The failure to recognize the human capability for semantic and pragmatic inferences, still unmatched by any computer program, can lead researchers to misclassify such intelligent inferences as reasoning errors.

In contrast to the probability instruction in the Linda problem, Inhelder and Piaget asked children whether there are "more flowers or more primulas." "More" refers directly to numerosity and does not leave open as many possible interpretations as the semantically ambiguous term "probability." Similarly, asking for "frequency" judgments in the Linda problem avoids the ambiguity of "probability" by narrowing down the spectrum of possible interpretations. This is a crucial reason why frequency representations can make conjunction effects disappear (Hertwig & Gigerenzer 1999; for another key reason, namely, the response mode, see Hertwig & Chase 1998).

In sum, Stanovich & West present the Linda problem as support for their thesis that higher cognitive ability underlies correct judgments in reasoning tasks. Whether applied to research by Inhelder and Piaget or to research within the tradition of the heuristics-and-biases program, however, the notion of "cognitive ability" gives rise to paradoxical conclusions. Rather than resort to ill-specified terms and vague dichotomies, we need to analyze cognitive processes – for instance, application of Gricean norms of conversation to the task of interpreting semantically ambiguous terms – which underlie people's understanding of the ambiguous reasoning tasks. Otherwise, *intelligent* inferences will continue to be mistaken for reasoning errors.

Why the analyses of cognitive processes matter

Ulrich Hoffrage

Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, 14195 Berlin, Germany **hoffrage@mpib-berlin.mpg.de**
www.mbip-berlin.mpg.de/ABC/Staff/hoffrage/home-d.htm

Abstract: Stanovich & West analyze individual differences with respect to response output (e.g., participants' numerical estimates). They do not analyze the underlying cognitive processes that led to the outputs; they thereby probably misclassify some non-normative responses as normative. Using base rate neglect and overconfidence as examples, I demonstrate the advantages of analyzing cognitive processes further.

Stanovich & West's (S&W's) criterion for classifying a response to a Bayesian inference task as normatively correct can be used to illustrate a major problem with their target article. In the cab problem, for which a posterior probability estimate of .41 is considered the normatively correct answer, "responses over 70% were classified as reliant on indicant information, responses between 30% and 70% as Bayesian, and responses of less than 30% as reliant on indicant information" (sect. 4.5, par. 5). Because this classification is only based on participants' outcome response outputs (here, numerical estimates), it can give rise to misclassifications. In particular, nonnormative response strategies that accidentally lead to a numerical estimate between 30% and 70% are classified as normative responses. For example, some participants may have subtracted the false alarm rate (20%) from the hit rate (80%), a strategy known as ΔR. This has been discussed as the correct strategy for estimating the covariation between two dichotomous variables (McKenzie 1994), and as a model of how people assess casual strength (Cheng & Novick 1992). In studies with medical students and physicians, Hoffrage and Gigerenzer (1999) found that ΔR was the second most frequently used strategy when information was presented in terms of probabilities, even more prevalent than the Bayesian strategy. In S&W's analyses of the cab problem, where ΔR would have resulted in a numerical estimate of 60%, these responses would have erroneously been classified as normatively correct. Similarly, in their AIDS problem, providing the base rate of the disease, or the joint occurrence of disease and positive test results as the numerical estimate would also have been classified as normatively correct. We avoided this type of misclassification by testing multiple strategies (e.g., Bayesian, hit rate, ΔR). The classification was based on two criteria: first, we compared a particular numerical estimate with the outcomes of various strategies; second, we analyzed participants' reasoning processes by inspecting their "write aloud" protocols and by conducting interviews after the questionnaires had been completed (Gigerenzer & Hoffrage 1995, p. 692). S&W, in contrast, seem to be interested neither in testing multiple strategies nor in analyzing the cognitive processes underlying participants' responses. Their misclassifications may hence have distorted the correlations they report; this is a matter for concern, as these correlations are crucial to their four explanations for the normative/descriptive gap.

How could S&W have benefited from going on to analyzing underlying processes? Consider their first explanation: performance error. The strongest form of the performance error hypothesis (i.e., that *all* deviations can be explained by performance errors) "has the implication that there should be virtually no correlations among non-normative processing biases across tasks . . . because error variances should be uncorrelated" (sect. 2, para. 4). This is not necessarily true, for differences in motivation, attention, or fatigue may affect performance. Because it is plausible that an attentive person tends to do well on all tasks, whereas an inattentive person will probably do badly on all tasks, one could argue that the performance error hypothesis likewise implies substantial correlations among tasks. As a consequence, the positive correlations reported by S&W are also compatible with the hypothesis that most of the errors were performance errors, and thus their conclusion "performance errors are a minor factor in the gap" (Abstract) is not justified.

Analyzing participants' processes provides a better test of the performance error hypothesis. For example, in our studies on Bayesian inferences (Gigerenzer & Hoffrage 1995; Hoffrage & Gigerenzer 1999) we identified cases where participants performed a Bayesian computation, but made a calculation error due to carelessness rather than to computational limitations. As these cases were very rare, we also failed to find support for the strong form of the performance error hypothesis. The difference is that S&W's conclusion is based on a questionable premise (performance errors entail no inter-task correlations), whereas analyzing processes provides tangible evidence for the claim that performance errors seem to play a negligible role.

Let me now turn to S&W's "wrong norm" and "alternative task construal" explanations. As they repeatedly pointed out, the appropriate normative solution for many problems is controversial. This is because there is not only a single solution, but many normatively correct solutions, depending on what assumptions are made. The authors' analyses of individual differences do not offer unequivocal help in settling such controversies in the rationality debate. Even if it should turn out that those participants who make the same assumptions about a particular task as the researcher also tend to be those with higher SAT scores, this would not necessarily invalidate the assumptions made by the participants with lower SAT scores. On the other hand, a zero correlation, as S&W reported for the overconfidence effect, does not provide strong evidence that participants constructed the task differently or that a "wrong norm" has been applied.

What could an analysis of cognitive processes contribute to evaluating S&W's third and fourth explanations, "wrong norm" and "alternative task construal"? The theory of probabilistic mental models (PMMs), for instance, postulates the cognitive processes underlying confidence judgments in the accuracy of single responses, and those underlying estimates of the number of correct responses in a series of questions (Gigerenzer et al. 1991). According to the theory, participants construct a PMM for each of the two tasks. Each PMM is based on participants' assumption that the items have been representatively drawn from a reference class. The theory further posits that the reference classes are different for the two PMMs. It follows that if the assumption for one PMM is justified, then the assumption for the other is violated, and vice versa. Thus, a participant cannot adhere to both norms – well-calibrated confidence judgments and accurate frequency estimates – at the same time, that is, for the same set of items. This is indeed what we found for a vast majority of participants. The example illustrates the interplay between participants' assumptions and researchers' norms. If for the same participant obeying one norm implies violating the other, it is to no avail to relate this participant's responses to his SAT score. Whether assumptions are justified and the experimenter's norms are met often has little to do with cognitive ability, but with the experimenter's construal of the material, thus bringing into question the external validity of laboratory findings.

In summary, it would be useful to test multiple strategies and analyze cognitive processes to avoid misclassifying responses and to better justify some conclusions drawn in the target article, such as the negligible role of performance errors. Whereas correlations based on outputs have only minor implications for the rationality debate, the identification and the modeling of processes is more likely to help us understand when and why people respond according to the experimenter's norms. Identifying processes is not only of theoretical but also practical importance. Although discovering that those participants who fail to respond consistently with a norm are also those with lower SAT scores may be of little use for education, having identified the processes underlying their responses may provide a clue to improving their reasoning.

Situational constraints on normative reasoning

Earl Hunt

Department of Psychology, University of Washington, Seattle, WA 98195-1525 **ehunt@u.washington.edu**

Abstract: Stanovich & West claim that the positive correlation between reasoning tasks negates the view that errors in reasoning are due to failures in information processing. This is not correct. They conjecture that errors in reasoning are associated with conflicts between intentional and associative reasoning. This interesting proposition suggests studies relating situational characteristics to the quality of human reasoning.

Stanovich & West's (S&W's) target article contains a number of important ideas. Possibly the most important is the demonstration (once again!) that individual differences in performance are not "error variance." People are not robots; attempts to make them behave as if they were passive carriers of information are fundamentally mistaken. This is the latest of several papers in which these authors have made this point with respect to a variety of "everyday reasoning" tasks. I will comment on three points in their paper. As is traditional in psychology, which is a gloomier science than economics, I will offer the bad news first, followed by the good news and the very good news.

The bad news has to do with what Stanovich & West (S&W) term the Panglossian explanation of non-normative responding in a variety of intellectual tasks. According to the Panglossian explanation, people are attempting to apply the normative model, but occasionally make errors due to information processing limitations. S&W challenge this explanation, on the grounds that errors in these tasks are positively correlated, and positively correlated with scores on conventional intelligence tests. The empirical finding will not surprise anyone familiar with either the psychometric literature or attempts to combine the psychometric literature with the literature from cognitive psychology (Jensen 1998; Macintosh 1998). The correlations in their Table 1 are slightly higher than, but in the range of, the remark that all tests of cognition have a correlation of about .3 with each other (Hunt 1980). But what does it mean?

The explanation cannot be that "general intelligence is involved" because this merely restates the phenomenon. The Panglossian explanation that failures of normative reasoning are due to random errors is compatible with the assumption that the failure is due to a random error, combined with the assumption that individuals with low general intelligence are more prone to make information processing errors than others. There are influential theorists who claim that this is exactly the case (e.g., Jensen 1998). S&W cannot knock down this particular Panglossian assertion with the evidence they offer.

Now to the good news. S&W offer two interesting possible hypotheses for the systematicity of non-normative performance. One is that good performance requires a high capacity working memory/and-or attention-control. In practice, the two are hard to distinguish. They combine this with the suggestion that errors in normative reasoning reflect a failure to inhibit "automatic" processing (e.g, automatic memory arousal) in order to conduct controlled, working-memory intensive processing. They point out that certain tasks that do not show systematic non-normative performance seem to be associated more with overlearned associations in memory than with controlled reasoning. This is consistent with their hypothesis. This idea is a very good one. It is not inconsistent with the Panglossian assertion that people with low intelligence are making more random errors, because the correlations cited are low enough so that there is plenty of room for the operation of several factors that lead to non-normative reasoning. This is a fruitful research hypothesis that deserves exploration.

And in that vein, I come to the very good news. The S&W hypothesis suggests exploration of situational constraints on normative reasoning. People should resort to non-normative reasoning in precisely those situations that overload working memory, or that emphasize automatic processing. There are a number of experimental paradigms that do precisely this. Dual tasks are the most obvious, but there are others. Intra-individual variabilities, such as fatigue or mild intoxication, could also be studied. The S&W hypothesis would be to show increases in non-normative reasoning in tasks that are supposed to be sensitive to working memory and attention, in experimental situations where the working memory-attention functions are overloaded. Tasks that are associated with controlled processing should show deficiencies, tasks that are not associated with controlled processing should not show such defects.

Understanding/acceptance and adaptation: Is the non-normative thinking mode adaptive?

Jerwen Jou

Department of Psychology and Anthroplogy, University of Texas-Pan American, Edinburg, TX 78539-2999 **jjou@panam.edu**
w3.panam.edu/~jjou

Abstract: The finding of a correlation between normative responses to judgment and reasoning questions and cognitive capacity measures (SAT score) suggests that the cause of the non-normative responses is computational in nature. This actually is consistent with the rational competence view. The implications of this finding for the adaptation view of cognition are discussed.

The Scholastic Assessment Test score and normative response in judgment and reasoning tasks are similar in that both use superficially plausible distractors (in multiple choices) to "entrap" the test takers. When the test taker is not reflective, engaged, or intelligent enough, he is likely to be lured into the "wrong" answer. Hence the correlation between the SAT score and the normative response is not too surprising. Logic reasoning (analytic ability) is actually one of the tests in GRE. It would be interesting to calculate the correlation between this part of the GRE and the normative response. I predict that the correlation will be much higher than that between SAT and normative responses. Stanovich & West (S&W) have made an important contribution in expanding Slovic and Tversky's (1974) original notion of understanding/acceptance by demonstrating a correlation between the traditional measure of cognitive capacity and the normative response to the judgment and decision questions. As they note, this is embarrassing to the reject-the-norm-application theorists and increases the confidence in the appropriateness of the normative models.

This finding implies that more intelligent people have larger information-processing capacity, enabling them to understand and hence accept the axiom more readily. Although this finding provides a strong rebuttal to the alternative construal view ("the subjects are right and the experimenters are wrong"), it makes the irrationality position less interesting than it was. Before the correlation between the traditional measure of intelligence and the normative response was known, the bias in judgment and decision making could be attributed to irrationality in the human cognitive system. Now, the biased response is reduced to no more than insufficient intelligence or limited information-computation capacity. In this sense, the finding is also embarrassing to the Meliorists (the irrationality camp) because it supports the computational-limitation rather than the irrational competence view of heuristics. Indeed, it is always possible to devise a problem which demands more computational capacity than is available to subjects and to elicit systematic errors. As S&W note, the errors caused by computational limitations, unlike the errors caused by lapses of attention and incidental memory failures, are nonrandom. Computational limitations in the human brain, as in the computer, are always a real constraint (Tsotsos 1991). The bad news from S&W's studies for the irrationality camp is that these judgment problems are just another way of measuring intelligence. Thus, what the bias and heuristics results have so far shown is no more than the gap between the intelligence of elite cognitive psychologists and that of average people. Does the finding imply that more intelligent people are also more rational, or that the more intelligent people simply have larger computational capacity?

S&W also suggest that the goal of rationality conflicts, under many circumstances, with the goals of the genes, which is adaptation. However, they point out near the end of the target article, that because of the high technology demands of modern society, the adaptive genes which are the products of evolution may no longer be adaptive. Instead, rational and analytic computational ability may become adaptive in the high tech modern world. If contextualized and socialized thinking is the product of evolution of the previous millennium, then, at the end of the next millennium, we will have average people who will all be thinking and making decisions as Tversky and Kahneman do (but if the evolution of the human cognitive apparatus always lags behind the evolution of the society, then these people will again be characterized by future cognitive psychologists as biased and irrational).

The adaptative view is actually also Anderson's (1991); according to whom the way our mind works reflects the structure of the environment. Since the high tech world is a very recent occurrence, we cannot expect average people to change their reasoning system overnight to meet the demands of this new high tech world. If their way of thinking is no longer adaptive in the modern world, it is not because their cognitive system is not adaptive. It is simply because the world is changing too fast. Hence in defending the normative model with the understanding/acceptance evidence, S&W actually help support the view that average people's irrational response to the judgment and reasoning problems was adaptive in their ancestors' world, if not in the modern one.

Nor is there evidence that the scenarios in the problems used by the judgment and decision researchers are representative of the average or typical ecological conditions. An abstract probability of occurrence for an event has little meaning in the real world unless one experiences that event many times over a long period of time (Gigerenzer 1994; Jou & Shanteau 1995). Some of the "irrational" thinking and behaviors may be the product of past adaptation and are accordingly adaptive under the majority of the real world's circumstances. The "sunk cost fallacy" is thought to be the product of a "waste not" or "persevere" rule. It is only when the rule is overgeneralized that it becomes counterproductive (Arkes & Ayton 1999).

Some violations of the transitivity rule may likewise be adaptive. For example, Brand X cola (60 cents) may be preferred over Brand Y cola (55 cents) for its higher quality. Brand Y cola is in turn preferred over Brand Z cola (50 cents), likewise because of its higher quality. But when Brand X and Brand Z are compared, Brand Z may be preferred over Brand X (Einhorn 1982). When adjacent brands are compared, the price difference is only 5 cents. So, quality determines the choice. When Brand X and Brand Z are compared, the price difference is larger. Consequently, the price factor dominates the choice. This intransitivity, although irrational, may have resulted from considering multiple dimensions when comparing different members of the series; this may be more adaptive than considering a single dimension.

Normative responses to some of the judgment problems were shown to be positively correlated with cognitive capacity, which is consistent with the understanding/acceptance notion. However, normative responses to some other problems (such as the Cab Problem and the AIDS Problem) are negatively correlated with cognitive capacity. Responses to still other problems (such as the Disease Problem) showed little correlation with scholastic aptitude (Stanovich & West; 1998b; 1999). Instances of these reversal and neutral cases seem to be more than just random irregularities. These cases detract from the value of the understanding/acceptance theory. S&W explain these inconsistencies by suggesting that a problem can cue either an analytic (or rational) thinking mode or a contextualized (i.e., intuitive) thinking mode. Depending on which thinking mode is cued more strongly, a response corresponding to the dominant mode will be made. Their explanation is post hoc. In addition, they do not specify what aspects or features of a problem will cue the normative thinking, and what will cue the non-normative thinking mode.

Until we can clearly identify and define the factors or features of the problems that control the response patterns of people of high and average scholastic aptitudes and that interact with the cognitive capacity, and until the inconsistencies in the pattern of correlations between the traditional scholastic aptitude measures and the normative responses can be reliably predicted (rather than explained post hoc), the understanding/acceptance theory remains rather opaque. However, Stanovich & West have started off in the right direction and opened up a new avenue to pursuit and possibly resolve this contentious issue of human rationality.

A psychological point of view: Violations of rational rules as a diagnostic of mental processes

Daniel Kahneman

Department of Psychology and Woodrow Wilson School of Public Affairs, Princeton University, Princeton, NJ 08544 **kahneman@princeton.edu**
www.princeton.edu/~psych/PsychSite/fac_Kahneman.html

Abstract: The target article focuses exclusively on System 2 and on *reasoning rationality:* the ability to reach valid conclusions from available information, as in the Wason task. The decision-theoretic concept of *coherence rationality* requires beliefs to be consistent, even when they are assessed one at a time. Judgment heuristics belong to System 1, and help explain the incoherence of intuitive beliefs.

As all its readers will be, I am grateful to the authors of this unusually thorough and erudite survey of the contentious debate on human rationality. I should start with a disclaimer: questioning human rationality is not my lifework. Contrary to a common perception, researchers working in the heuristics and biases (HB) mode are less interested in demonstrating irrationality than in understanding the psychology of intuitive judgment and choice. The purpose of studying deviations of human judgments and choices from standard rational models (e.g., Bayesian reasoning, expected utility theory) is to develop diagnostics for particular psychological processes. In the same spirit, hundreds of psychologists have studied Wason's selection task for the light it sheds on human logical reasoning, not because it threatens the rational agent model. As many of us have found, however, our tool – the analysis of errors in reasoning and in intuitive judgment – attracts vastly more attention than the results we obtain with it. This selective interest is reflected in the target article, which addresses failures of rationality but not their psychological causes.

The core of Stanovich & West's (S&W's) research program is the study of correlations between errors of thinking and standard measures of intellectual aptitude. Their main purpose is to assess the prescriptive value of principles of logic and probability, but some of their results are also of considerable psychological interest. For an example, consider the finding of a positive correlation between the SAT and susceptibility to a common error of judgment – overweighting the indicant probability in noncausal base-rate problems. I conclude from this finding that non-causal base-rate problems – unless they are made trivial by providing broad hints – are too difficult for almost everybody. The Bayesian way of combining new evidence with prior beliefs is not intuitive for anyone, and only exists in the repertoire of respondents who have had training in statistics. Faced with an intractable puzzle, some naïve respondents will interpret the problem as requiring a choice between the base-rate probability and the indicant probability. The choice of the indicant is more plausible than the alternative, and the evidence reported in the target article shows that this choice is made more often by the more intelligent respondents.

In the terms of the target article, a task will be too difficult if (1) System 1 favors an incorrect answer, and (2) System 2 is incapable of applying the correct rule, either because the rule is unknown or because the cues that would evoke it are absent. This conjunction of conditions is probably quite frequent. For example, my impression is that prediction by representativeness may well be universal – and therefore uncorrelated with intelligence. I formed this impression over years of observing my colleagues (and myself) enthusiastically making nonregressive predictions of a job candidate's academic future on the basis of one or two brilliant answers to questions raised in a talk. All of us know about regression to the mean, but the task of evaluating a job candidate rarely evokes that statistical fact. The correlational diagnostic that S&W have proposed works as intended only in the intermediate range of difficulty; it cannot identify the normatively appropriate solution unless some naïve respondents find that solution. If problems are too difficult, carefully devised hints to the relevance of base rates may be provided, allowing System 2 to guide some respondents to the correct answer. Because talented respondents will be most likely to detect these hints and to benefit from them, the correlation between the use of base rates and intelligence should now be positive.

I now turn to a more fundamental issue. As I understand it, the term "rationality" is used differently in the target article and in the rational-agent models of decision theory, economics, and political science. In decision theory, rationality is defined by the coherence of beliefs and preferences, not by the ability to reason correctly about immediately available information. I will use the terms coherence-rationality and reasoning-rationality to distinguish the decision-theoretic concept from the concept that S&W address. Performance in Wason's celebrated selection task is the prime example of a failure of reasoning-rationality. In contrast, the biases associated with judgmental heuristics are best viewed as failures of coherence-rationality. The coherence criterion should not be dismissed lightly. Whether or not it is psychologically plausible, the assumption that beliefs and preferences are coherent is essential to the dominant theoretical model of social science.

The distinction between reasoning-rationality and coherence-rationality is of immediate methodological relevance, because the two concepts are best studied in different research designs. For example, consider the famous Linda. A scholar interested in reasoning-rationality will be drawn to experiments that implicitly or explicitly require respondents to compare the probability (or the frequency) of two statements: "Linda is a bank teller" and "Linda is a bank teller and is active in the feminist movement." Although the target article lists many excuses for them, respondents who see both statements and still assert that the latter is more probable than the former display faulty reasoning. If a scholar wishes to study intuitive judgment, however, more will be learned by asking different respondents to judge the two critical outcomes (Kahneman & Tversky 1996). The robust finding of a conjunction effect in such between-subjects studies demonstrates systematic incoherence in people's beliefs – even if no particular individual can be described as incoherent.

For another example, consider a request to compare the probabilities that an individual will die within a week, or within a year. Von Mises to the contrary notwithstanding, most of us would say that anyone who does not answer this question in the obvious manner fails a test of reasoning-rationality (or, alternatively, is a philosopher). This is the within-subject version of the problem. The between-subjects tests of coherence is much stricter. It requires respondents to be disposed to produce the *same* judgments of probability, regardless of whether the questions about the week or the year are asked together or separately. Furthermore, coherence requires choices and beliefs to be immune to variations of framing and context. This is a lot to ask for, but an inability to pass between-subjects tests of coherence is indeed a significant flaw. Knowing rules and being able to apply them is good, but not sufficient, because much of life resembles a between-subjects experiment. Questions about preferences and beliefs arise one at a time, in variable frames and contexts, and without the information needed to apply relevant rules. A perfect reasoner whose judgments and choices are susceptible to framing and context will make many errors in the game of life.

The focus on reasoning-rationality in the target article fails to make contact with the study of heuristics of judgment. Tversky and I always thought of the heuristics and biases approach as a two-process theory (Chaiken & Trope 1999). We proposed that intuitive judgments of probability or frequency are based on "natural assessments" – that is, of similarity, causality, affective valence, or past frequency – which are effortless and largely automatic. In the terminology of the target article, judgment by heuristics is a manifestation of System 1. However, we also believed that System 2 can override intuitive judgments that conflict with a general rule – but only if the relevant rule is evoked. For example, we observed that trained statisticians easily avoided the conjunction fallacy when "bank teller" and "feminist bank teller" were to be judged in immediate succession. However, trained statisticians did no better than statistically naïve respondents in a between-S design, or even when the two critical outcomes were embedded in separate positions in a list of eight possible outcomes (Tversky & Kahneman 1983). The fact that statistically sophisticated respondents fail the eight-item test but pass the two-item version demonstrates that training in statistics (1) does not affect the intuitive beliefs produced by System 1, (2) does not eliminate prediction by representativeness, but (3) improves System 2 reasoning and its ability to detect cues to relevant rules.

If this analysis is correct, the positive correlation between performance in the Linda problem and the SAT should disappear if an eight-item list is used. Whether a within-S or a between-S design is appropriate depends on the investigator's purpose. Within-S designs are needed to study the process by which the deliberate application of rules overrides intuitive thinking (Kahneman & Tversky 1982). However, the opportunity to reason by rule is a confounding factor when the aim is to study heuristics, intuitive judgment, or the coherence-rationality of beliefs.

A recent development in the Panglossian view is an emphasis on 'heuristics that make us smart' (Gigerenzer, Todd & the ABC Research Group, 1999). The phrase suggests – without exactly saying – that these heuristics do not induce biases. But this cannot be true. The biases that heuristics induce are not constant errors – the type of error that a friendly scale makes in understating all weights. Heuristics, by definition, overweight some sources of relevant information and neglect others, thereby inducing *weighting biases.* The diagnostic of a weighting bias is a partial correlation: judgments are correlated with any variable that the heuristic overweights, with the true value held constant. Thus, the heuristic of judging the distance of mountains by the blur of their contours (aerial perspective) leads to an underestimation of distance on clear days and an overestimation of distance on hazy days. Similarly, the recognition heuristic (Gigerenzer, Todd & the ABC Research Group 1999) will necessarily induce a large partial correlation between familiarity and recognition-based judgments, with truth constant. All heuristics make us smart more often than not, and all heuristics – by mathematical necessity – induce weighting biases.

Whether heuristics make us differentially smart is an interesting question. The target article has focused on the relation between conventional measures of intelligence and System 2 functioning. However, System 1 may have its own kind of intuitive intelligence. For example, some people may have particularly nuanced and subtle representations of persons and of social categories. These people will make better judgments by representativeness than others, and may consequently achieve greater predictive accuracy than others – especially if everybody ignores base-rates. The successful research program described in the target article could be the beginning of a new wave of studies of individual differences in both intuitive and rule-governed thinking.

Is rationality really "bounded" by information processing constraints?

Paul A. Klaczynski

Department of Psychology, Western Carolina University, Cullowhee, NC 28723 **klaczynsk@wcu.edu**
www2.ceap.wcu.edu/%7EKlaczynski/

Abstract: Extremist views on normative rationality fail to address differences in responding owing to intellectual ability or epistemic self-regulation. Individual difference research thus raises serious questions concerning the scientific utility of universal rationality and universal irrationality theories. However, recent data indicate that computational capacity theories do not account adequately for within-subject variability in normative responding, memory-reasoning independence, and instances of ability-normative reasoning independence.

Stanovich & West (S&W) thoroughly debunk Panglossian perspectives on rationality, regardless of whether these derive from "reject the norm," "random performance error," or "alternative task construal" arguments. The message, that all people are not equally rational, is supported by considerable evidence of individual differences in normative responding. The same evidence seems to support explanations of between-subject variability that recognize the confluence of algorithmic-level limitations and intentional-level differences in epistemic self-regulation.

The Panglossians have lost the war, but so too have strict Meliorists. Just as systematic individual differences in normative responding dispel myths of universal rationality, they falsify claims from the universal irrationality camp. Can those who advocate computational limitations as the boundary between irrationality and perfect rationality claim victory? For reasons detailed below, I believe the answer is no.

1. Considerable leaps of faith are required to believe that SAT scores index capacity limitations, that computational imitations are domain-general impediments to reasoning, and that the reported correlations indicate causality. Even if these assumptions are accepted, normative responding on many tasks often fails to be constrained by these limitations, as the overconfidence and false consensus findings indicate. I would add data from three unpublished studies (*ns* 128–78): Outcome biases ($r = .04$), hindsight biases (.06), unrealistic optimism (−.07), using causal base rates (.14), self-contradictory reasoning (.05), associative reasoning (−.11), honoring sunk costs (−.06), and myside biases (−.05) are unrelated to ability. From a computational limitation standpoint, the "correct" norms for these tasks are immaterial: Those with more capacity should have performed differently from those with less.

2. Computational limitation explanations fail to explain phenomena such as task variability (Reyna & Brainerd 1995). On logically isomorphic problems, performance varies considerably within individuals, violating internal consistency criteria for normative rationality. On scientific reasoning problems that have no obvious relation to individuals' beliefs, "capacity" predicts performance. When, in contrast, the same problems contravene beliefs, both high and low ability subjects perform remarkably well. Likewise, when problem-content supports beliefs, reasoning is poor – regardless of ability. In within-subjects designs, this problem-to-problem vacillation between System 2 processing and System 1 processing is independent of ability (Klaczynski & Gordon 1996; Klaczynski & Robinson, in press). Such belief-biases not only violate the principle of descriptive invariance but cannot, without extraordinary efforts, be explained by bounded rationality theories.

3. Biological maturation invariably increases computational capacity. Consequently, if computational limitations are fundamental hindrances to normative responding, then such responding should increase with age. Although early Piagetian and information processing research supported this conjecture, research focusing on children's responses to heuristics and biases tasks paints a more complicated picture. The key finding for these endeavors – that *non-normative* responding increases with age – obviously cannot be explained by increases in capacity. Under certain conditions, framing effects (Reyna & Ellis 1994), base rate neglect (Jacobs & Potenza 1991), conjunction fallacies (Davidson 1995), denominator neglect (Brainerd 1981), and conditional reasoning fallacies (Klaczynski & Narasimham 1998) increase with age. In conjunction with age-related increases in logico-computational reasoning, these data indicate that development involves two progressions – one toward increased decontextualization, the other toward increased contextualization. It is with the latter progression that bounded rationality theorists should be concerned.

4. Most cognitive scientists agree with S&W that "working memory is the quintessential component of cognitive capacity," but there is reason to doubt the centrality of memory to reasoning. If working memory is the principle barrier to normative responding, then reasoning on tasks that appear to demand memory should depend on verbatim retention of problem information. This hypothesis has been treated axiomatically, despite repeated demonstrations of memory-reasoning *independence* (Reyna & Brainerd 1995).

The independence findings are explainable by assuming two dissociated memory systems – memory for gist and for verbatim details. Memory independence, memory interference, and expertise research show that theories of reasoning need not invoke notions of limitations in generic processing resources or in capacity (Reyna 1995). Nonetheless, most two-process theorists continue to rely on limitation arguments. If these are to remain viable, data challenging their fundamental tenets must be addressed rather than ignored.

None of this is to deny that there are limitations on the complexity of the mental operations people can perform. The point is instead that there are compelling empirical grounds for doubting that these limitations are the only, or even the most important, reason for the pervasiveness of non-normative responding and for individual differences in responding.

S&W propose that ability predicts performance when Systems 1 and 2 are brought into conflict. If this is correct, capacity theorists might argue that the "pull" for System 1 and System 2 processing is the same across subjects. Like high ability subjects, low ability subjects attempt System 2 reasoning but are overwhelmed by problem complexity and fall back on the default system. However, the mechanisms separating high and low ability subjects may have nothing to do with capacity. As S&W (sect. 6.1, para. 4) suggest, the "pull" toward the normative rules of System 2 may be stronger for higher ability subjects. If "the most important difference between [System 1 and System 2] is that they tend to lead to different task construals" (sect. 6, para. 4), then the principle difference between ability levels may be the representations upon which further operations are performed, as the selection task results suggest. These representational differences may lie in the degree of decontextualization (e.g., gist, verbatim) or in type of decontextualized representation (e.g., inductive, deductive).

In focusing on individual differences, S&W have dealt a damaging blow to extremists positions on rationality. They have thereby also provided some clues as to the processes distinguishing high from low ability individuals. Decontextualization, both a product and a producer of formal educational and economic success (Rogoff & Lave 1984; Schooler 1984), may rarely be demanded in our everyday affairs. Nevertheless, the stakes are sometimes quite high when such System 2 processing does not take place, when politicians, religious leaders, and scientists fail to reason independently from erroneous beliefs, physicians disregard statistical base rates in forming diagnoses, and students do so when assessing career opportunities.

When decontextualization failures are related to psychometric intelligence, we must not leap too quickly to capacity explanations. The need to evaluate other plausible explanations is great, for if decontextualization is a domain-general ability, and if that ability is determined, to some degree, by capacity limitations, then there may be little hope for educators who seek to prepare low ability students for the demands of an increasingly technological society.

Individual differences and Pearson's *r:* Rationality revealed?

Joachim Krueger

Department of Psychology, Brown University, Providence, RI 02912
joachimkrueger@brown.edu
www.brown.edu/Departments/Psychology/faculty/krueger.html

Abstract: Regardless of the clarity of the patterns they produce, individual differences in reasoning cannot validate norms of rationality. With improved reliability, these correlations will simply reveal which sorts of biases go together and which predict the intelligence of the decision maker. It seems necessary, therefore, to continue efforts to define rational thought independently of intelligence.

The heuristics-and-biases paradigm suggests that ordinary people often fail to think rationally. This view is supported by reliable differences between average or modal human judgments and relevant normative standards. Individual differences are typically banished to the error term of the test statistic, and thus much information is lost. Stanovich & West (S&W) bring individual differences back into focus. This is an important development because (ir)rational thinking presumably occurs within individuals, and some individuals reason more rationally than others.

Group-level analyses not only ignore systematic variations among people, they also work against the vindication of human judgment (Krueger 1998a). Significant discrepancies between predicted (i.e., normative) and average actual judgments signal the violation of a norm. Such violations can easily be detected even if only a minority of participants responds non-normatively. For example, significant minorities allow past investments to affect decisions about the future, show intransitive preferences, and conform to the false judgments of a planted majority. Klar and Giladi (1997) concluded that people believe that "everybody is better than average." On the average, individual group members were indeed liked more than the group itself, but there was no difference in the modal response. Statistical outliers can thus distort group-level analyses and bias inferences about general psychological processes.

The individual-differences approach is more conservative in drawing categorical conclusions regarding (ir)rationality. Instead, it reveals some interesting patterns. In S&W's (1998c) study, rational reasoning appears to transfer from one task to another. Correlations among individual differences in rational responding ranged from .12 to .36 ($M = .25$, judging from Tables 1 and 2). However, with the exception of the argument-evaluation task (23 items), reasoning was assessed by only 1 (e.g., outcome bias) to 8 (e.g., syllogistic reasoning) problems. Once reasoning abilities are measured more reliably with multiple-item scales, these correlations will probably increase and strengthen the argument that performance errors cannot explain norm violations. Increased reliability will also boost the correlations between normative reasoning and the psychometric g factor of intelligence (now $M = .23$, Tables 2 and 3). S&W's (1998c) data already show that these correlations increase with the number of problems used to measure reasoning ($r = .42$).

Improved measurement will raise two new questions. First, is there a g factor of rational thinking? If high intertask correlations remain when intelligence is controlled, it will be possible to postulate the existence of trait of rationality. Second, will rationality remain separate from intelligence? As correlations between rationality and intelligence increase, the temptation to subsume the former under the latter will also increase. S&W entertain the idea that being rational is just one way of being smart. They suggest that "examining individual differences may actually reinforce confidence in the appropriateness of the normative models applied to problems in the heuristics and biases literature." Some biases (overprojection and overconfidence), however, are unrelated to intelligence. This state of affairs is unlikely to change with improved measurement. Another bias (ignoring noncausal base rates) is even negatively correlated with intelligence. These exceptions to the "positive manifold" remain inexplicable if g is the only benchmark for rationality.

The attempt to justify normative models of rationality with correlations between rational responding and g undermines efforts to devise and deploy independent criteria of rationality. The positive manifold among measures of rationality and intelligence simply suggests that good traits go together, just as bad traits do. It can be shown that g is a factor contributing to rational judgment, but this does not say much about the quality of the normative models themselves. Correlations between rational reasoning and other desirable person characteristics share this limitation. As noted by S&W, Block and Funder (1986) found that well-adjusted adolescents are most generous in attributing a person's role-conferred success to that person's disposition. Although this finding nicely demonstrates that the fundamental attribution error does not necessarily predict other negative person characteristics, it cannot speak to the normative adequacy of the attributions themselves.

Another provocative idea of S&W is that normative responses are "not prescriptive for those with lower cognitive capacity." This conclusion conflicts with the claim that correlations between normative responding and g support the general validity of the norms. If the standard norms only apply to the bright, what norms are the dim-witted to be held to? How bright does a person have to be to be held to conventional standards of logical and statistical reasoning?

Criteria for normative reasoning must be found and justified independent of g. Most normative models are derived from some algebraic, probabilistic, or logical calculus. Their judgments are rational because they avoid contradictions, and not because they seem reasonable to well-educated and well-intentioned decision-makers (Dawes 1998). Sometimes, models that used to appear normative turn out to be flawed because contradictions are discovered. New models, which avoid these contradictions, then replace the old models. Some of these new models may improve predictions of human performance (as, for example, in the case of social projection; see Krueger 1998b), but they are not chosen for that reason.

What about motivation?

Anton Kühberger
Department of Psychology, University of Salzburg, A-5020 Salzburg, Austria
anton.kuehberger@sbg.ac.at
www.sbg.ac.at/psy/people/kuehberger.htm

Abstract: In their use of correlations as a means to distinguish between different views on the normative/descriptive gap, Stanovich & West discuss the competence component but neglect the activation-utilization component of performance. Different degrees of motivation may introduce systematic variation that is confounded with the variation explained by cognitive capacity.

In their discussion of the normative/descriptive gap, Stanovich & West (S&W) distinguish transitory performance errors from nontransitory computational limitations. They report significant correlations between tasks and conclude that this renders the performance-error view unlikely while being (weak) evidence for the computational-limitations view. However, their argument neglects an important aspect: being higher in cognitive capacity is not equivalent to investing more cognitive capacity. That is, besides measuring cognitive capacity, we must also measure the motivation to invest cognitive effort for a specific task at hand. Consider the case of stable cognitive capacity (as is usually assumed), but varying effort between problems (due to motivation, interest, situation, method, etc.). This blurs the distinction between transitory and nontransitory factors. That is, although cognitive capacity is stable, computational limitations may not be stable, owing to different degrees of cognitive effort. Thus, different degrees of cognitive effort may introduce systematic variation that is confounded with the variation explained by cognitive capacity. Consequently, both arguments cannot be made: that significant cross-task correlations speak against the performance errors view, and that significant correlations between performance and cognitive capacity indicate that the normative/descriptive gap is due to computational limitations.

The question of actual motivation is also important for the discussion of System 1 and System 2 processing. One prediction would be that individuals who are more highly motivated tend to do more System 2 processing. This is the essence of the effort-model of decision making: more cognitive effort leads to better decisions (in the sense that decisions do better conform to a normative model; Smith & Walker 1993). The intuition seems to be that participants will calculate more, think harder, or somehow see the appeal of axioms when they are faced with larger stakes. For higher motivation to change decision strategies and to increase performance, (1) one must believe that one's current decision strategy is insufficient in terms of desired accuracy; (2) a better strategy must be available; and (3) one must believe that one is capable of executing the new, more rational strategy (Payne et al. 1992). It is plausible that all three preconditions will be correlated with measures of cognitive capacity. In addition to stable dispositions, varying motivation (owing to different tasks and situations) may influence processing systems. Whether people activate System 1 or System 2 processing may depend on what people feel to be appropriate processing for the task at hand. For instance, in face-to-face interaction, contextual thinking could possibly be considered more appropriate than decontextualized thinking. Or, as implied by S&W, in a framing task people may be less likely to show a framing effect under high motivation, while a high degree of motivation may be detrimental to performance in tasks that subjects find attractive and that require heuristic strategies (McGraw 1978).

In summary, individual difference data provide a useful basis for deepening our understanding of the normative/descriptive gap in the judgment and decision making literature. It is necessary, however, to make a distinction between competence and activation-utilization. As Overton and Newman (1982) argue, two distinct components are required for a complete psychological theory. One is a competence component that is an idealized model of an individual's abstract knowledge in a given domain (as discussed in the target article). The activation-utilization component encompasses the psychological procedures and situational factors that determine the manifestation of the competence (which is not sufficiently discussed in the target article).

g and Darwinian algorithms

Kevin MacDonald[a] and David Geary[b]
[a]*Department of Psychology, California State University-Long Beach, Long Beach, CA 90840-0901;* [b]*Department of Psychology, University of Missouri, Columbia, MO 65211.* **kmacd@csulb.edu csulb.edu~/~kmacd/**
gearyd@missouri.edu www.missouri.edu/~psycorie/

Abstract: Stanovich & West's assumption of discrete System 1 and System 2 mechanisms is questionable. System 2 can be understood as emerging from individuals who score high on several normally distributed cognitive mechanisms supporting System 1. Cognitions ascribed to System 1 and System 2 appear to be directed toward the same evolutionary significant goals, and thus are likely to have emerged from the same selection pressures.

In demonstrating that individuals who are high in *g* (System 2) are able to inhibit the operation of Darwinian algorithms (System 1) and thereby engage in decontextualized and abstract reasoning, Stanovich & West (S&W) have made an important contribution and provided a corrective to the views of many evolutionary psychologists. These psychologists downplay the importance of the domain-general abilities assessed by *g* and even question their existence (e.g., Tooby & Cosmides 1992). Moreover, many psychologists, including S&W, have not fully appreciated how the pursuit of goal states might allow for the evolution of domain-general mechanisms (MacDonald 1991). People able to devise and imitate social strategies and learn about unforeseeable contingencies in a manner that is largely free of context would be at an advantage in achieving evolved goal states such as social status and mating in complex, nonrecurring environments. The key is to understand the relation between goals states, Darwinian algorithms, and *g*.

Goal states are reflected in the emotional and motivational aspects of behavior as these are related to the pursuit of personally significant ends (Campos et al. 1989). The Five Factor Model of personality captures individual differences in the relative importance of the psychological rewards associated with the attainment of evolutionarily significant goals, including individual differences in the salience of psychological rewards related to successful risk-taking in pursuit of resources, sexual gratification, and social status (MacDonald 1991; 1995; 1998). We have no doubt that the evolution of heuristics aimed at solving recurrent problems of our evolutionary past is the best strategy in static environments that present recurrent problems and that many of these heuristics are still relevant today (Caporael 1997; Tooby & Cosmides 1992). However, an evolutionary advantage for attaining evolved goals such as social status in complex, rapidly changing, nonrecurring environments would be achieved by domain-general mechanisms able to: (1) abstract general principles independent of context; (2) learn nonrecurrent contingencies quickly and efficiently, and, via a large working memory; (3) manage several concurrent goals.

This perspective is compatible with that of Potts (1998), who argues that the environments of human evolution were so varied that humans have been selected for their ability to make adaptive responses in novel environments. And it is compatible with the finding that intelligence is linked to the attainment of social status in modern and traditional societies (Gottfredson 1997; Lynn 1997). Moreover, the empirical literature on *g* and the associated domain-general learning ability indicates that Darwinian algorithms do not provide a sufficient explanation for many aspects of human learning and cognition (Jensen 1998), as demonstrated by S&W.

We conjecture that the cognitions of individuals high in *g* (System 2) are directed toward the same goals and achieving the same evolved motive dispositions as individuals dominated by System 1 heuristics. In addition to the psychological goals mentioned above, these goals include achieving some degree of control over the social, biological (e.g., food), and physical (e.g., territory) resources that support survival and reproduction (Geary 1989). The associated System 1 heuristics reflect Darwinian algorithms in the domains of folk psychology, folk biology, and intuitive physics. System 2 processes, or *g*, would also be focused on the same goals but at a higher-level of abstraction. The formal disciplines in the humanities and the social sciences, for instance, are an abstraction of folk psychology. The biological sciences are focused on the same basic goals as folk biological knowledge, that is, learning about other species. The goals of the physical sciences are the same as intuitive physics, that is, learning about the inorganic world. In this view, individuals of high intelligence (System 2) are focused on many of the same issues as individuals whose behavior is dominated by System 1. The difference is in the level of abstraction and decontextualization, such as the ability to deduce and make explicit the implicit rules that comprise System 1, and the reliance on scientific rules of evidence to override System 1 biases and infer more accurate patterns of relationships.

Not only are the goals and motive dispositions the same for System 1 and System 2; we propose that the underlying cognitive mechanisms are as well. In other words, the assumption of a discontinuity between System 1 and System 2 mechanisms is questionable. The general framework of evolutionary psychology is that evolved systems are species-typical universals, which conflicts with the finding that only a small number of people consistently evidence System 2 abilities. Moreover, it is not clear why individual differences in *g* are normally distributed rather than bimodally distributed. A normal distribution of *g* might imply that System 2 abilities are potentially evident in all people, but expression is contingent on task complexity. Alternatively, if there is a continuity between System 1 and System 2 mechanisms, then a normal distribution of *g* would be expected and the abstraction and decontextualization features ascribed to System 2 would also be important features of System 1.

Indeed, some degree of abstraction and decontextualization is necessary for the operation of System 1 heuristics. The social heuristics that are encapsulated in System 1 operate in much the same way across people and across different real-world contexts. The use of System 1 heuristics across people and contexts implies some degree of abstraction and decontextualization. If this were not the case, then heuristics would need to be learned anew with each variation in social exchange and across different people. In fact, the study of System 1 heuristics involves some degree of abstraction; to solve these tasks people are presented with a written vignette that describes some form of social problem solving, which, in turn, requires people to construct an abstracted, mental model of this social problem. Thus, the abstraction and decontextualization ascribed to System 2 is also a feature of System 1, although in less obvious ways.

As noted by S&W, individual differences in working memory underlie much of the variation in *g* (Carpenter et al. 1990). Individual differences in working memory are more or less normally distributed and involve at least several components that are presumably features of System 1. For instance, working memory engages the articulatory and visual systems, as well as central control mechanisms, that support language and other evolved processes (Baddeley 1994). In other words, working memory (and by inference *g*) must have its roots, so to speak, in more elementary, System 1, mechanisms. Carpenter et al.'s (1990) analyses of individual differences in *g* suggest that highly intelligent (System 2) and less intelligent (System 1) adults differ in the ability to manage several concurrent goals and to induce abstract rules during problem solving. These, however, represent continuous and nondiscrete skills. In short, many of the cognitive features of System 2 are not qualitatively distinct from those supporting System 1. Rather, System 2 emerges from individual differences in the mechanisms supporting System 1. The performance of individuals at the high end on several mechanisms of System 1, such as abstraction and goal management, creates the illusion of a qualitatively distinct System 2.

Thus, there is no need to posit that System 1 reflects gene-level selection and System 2 reflects individual-level selection. If the mechanisms supporting System 1 are the same as those supporting System 2, then the associated selection pressures are likely to have been the same. Moreover, individual-level selection is a much more likely candidate for conceptualizing the evolution of the systems discussed here. Selection at the level of the genes implies competition between different genes within the genome (e.g., via meiotic drive) – not likely to be relevant here.

In this view, the ability to learn in modern society is the direct result of the selection pressures – such as social competition – that resulted in gradual improvements in *g* during the course of human evolution. Individuals with high *g* direct these resources toward the same basic goals – including the affective goals linked to personality theory (MacDonald 1998) – as individuals dominated by System 1, but do so fairly independent of context. Modern society does not present any unique difficulties, but reflects the gradual accumulation of social (e.g., laws) and physical (e.g., agriculture, jets) tools that have emerged from the struggle for the control of social, biological, and physical resources (Geary 1998). Intelligence results in (and from) and abstraction and de-contextualization of more basic forms of competition; the result is a complex, evolutionarily novel world where achievement of the evolved motive disposition of social status seeking, for instance, is increasingly linked to being high on *g* (Gottfredson 1997).

On the meaning and function of normative analysis: Conceptual blur in the rationality debate?

David R. Mandel

Department of Psychology, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, United Kingdom **d.r.mandel@herts.ac.uk**
www.psy.herts.ac.uk/pub/d.r.mandel/

Abstract: The rationality debate centers on the meaning of deviations of decision makers' responses from the predictions/prescriptions of normative models. But for the debate to have significance, the meaning and functions of normative analysis must be clear. Presently, they are not, and the debate's persistence owes much to conceptual blur. An attempt is made here to clarify the concept of normative analysis.

Questions about the rationality of human judgment and deliberative human behavior are nowadays largely the focus of decision science, which is often said to rest upon a tripod of *descriptive, normative,* and *prescriptive* analyses. As Stanovich & West (S&W) have made clear, the rationality debate certainly has much to do with the tension between descriptive and normative accounts. But how stable is the tripod of decision science? And how much of the rationality debate is really a product of conceptual blur?

Let me begin by asking a simple question: *What defines a normative analysis?* If normative analysis is so vital for assessing rationality, then the answer to this question should be straightforward – but it is not. Normative analysis often relies on integrating one or more axioms or theorems of a formal system and examining the implications for judgment or decision making (Bell et al. 1988b). For instance, the much discussed conjunction rule is a normative principle that derives from probability theory: Probabilities themselves never violate the rule, yet people may or may not reason in accordance with it – that is an empirical issue.

Many other normative principles, such as transitivity, independence, and descriptive or procedural invariance, fit this definition, but others do not. The phi coefficient and the delta rule, often treated as normative models of contingency judgment and causal induction, respectively, are two examples. These information integration models achieve particular information processing objectives often valued by scientists but, in spite of their quantitative nature, they do not reflect properties of formal systems (Mandel & Lehman 1998). Other principles, such as Grice's maxims of conversation, which are often described as normative (Hilton 1995), seem utterly different again. So, strictly in terms of defining what is normative, it seems that decision scientists have been comparing apples with oranges.

Let me ask another question: *What is the primary function of normative analysis?* Clearly, many researchers use normative models as benchmarks to evaluate the "goodness" of people's judgments and decisions. From this Meliorist perspective, as S&W call it, rationality is seen as an inverse function of the deviation of the decision maker's response from the normative response. Hence normative analysis often serves a prescriptive function, despite the ostensible separation of normative and prescriptive foci in decision science. In contrast, normative analysis is sometimes used to generate *deductively* "first cut" descriptive accounts of judgment and decision making that are to be modified if necessary in light of later inductive analyses (Bell et al. 1988b). Expected utility theory, for instance, was initially used as a descriptive account of microeconomic behavior – not as a prescriptive benchmark for evaluating choice. Normative analysis may accordingly also serve a descriptive function, in spite of the ostensible separation of descriptive and normative foci in decision science.

Let us examine the prescriptive function often ascribed to normative analysis in greater detail, for it is that function which has influenced so many of the claims that decision scientists have made about rationality. According to Kleindorfer et al. (1993, p. 177), normative analysis focuses on "how decision makers *should ideally perform*" an activity. More dramatically, Bell et al. (1988b, p. 16) have claimed that "normative theory has something to do with how idealized, rational, super-intelligent people should think and should act." Normative analysis is often contrasted with prescriptive analysis, which is usually said to be geared toward examining what *real* people ought to do given their real-world constraints and cognitive limitations (or how decision scientists might aid real decision makers).

The notion of the *ideal* decision maker is, I believe, the result of a conceptual confusion. As noted earlier, many normative theories rely on axioms of formal systems that are used deductively to generate "first cut" descriptive accounts of human decision making. This approach tends to ask the following type of question: Do people judge or decide as some relevant set of abstract entities behave? For example, if probabilities behave according to principles *x, y,* and *z* in probability theory, do humans conform to principles *x, y,* and *z* in the process of judging probabilities? Normative analysis accordingly examines how real decision makers *would* perform if they were like idealized (i.e., abstract) entities of a formal system – not how they *should* perform if they are ideal (i.e., super-intelligent) decision makers. Probabilities, sets, and numbers are not super-intelligent entities – indeed, they have no intelligence – and it is unclear why adaptive decision makers *should* follow the same rules as they do.

Unfortunately, the preceding distinction seems to have been lost in the rationality debate and the erroneous conception of the ideal decision maker has often been adopted. If consistency and coherence are hallmarks of rationality, then Panglossians and Meliorists alike should question the rationality of some fundamental practices in decision science, for, at present, normative theories, on one hand, are said to describe what idealized decision makers should do and, on the other hand, they are used to evaluate real decision makers who are subject to myriad real-world constraints. This formula simply does not add up. Similarly, the use of the term *normative rationality* by S&W to describe instrumental rationality at the level of the individual is, I believe, confusing, because it has yet to be demonstrated that normative responses optimize individuals' multiple concerns.

Finally, let me be clear that, in my view, this critique is not a Panglossian perspective. I have not offered another reason why people might not be as irrational as some Meliorists would have us believe. Indeed, I have not commented at all on whether people are rational or irrational. Rather, I have questioned the conceptual rigor that underlies the overall debate. The meaning and function of normative analysis and the relation of normative analysis to descriptive and prescriptive concerns needs to be thought through more clearly in the future if the rationality debate is to optimize its significance.

Dilemmas of rationality

K. I. Manktelow

Division of Psychology, University of Wolverhampton, Wolverhampton WV1 1SB, United Kingdom **k.i.manktelow@wlv.ac.uk**

Abstract: This commentary focuses on the implications of practical reasoning research for the view of rationality in Stanovich & West's target article. Practical reasoning does not correlate with intelligence or other reasoning tasks. Explanation in decision making terms raises the issue of dilemmas, making it hard to specify the correct norm, when an action can satisfy or conflict with two equally justifiable goals.

Stanovich & West's (S&W's) research program will prove immensely valuable and influential to the psychology of thinking and to our debate about human rationality. Their central message, to look beyond the modal response and consider the range of performance on tests of reasoning, judgment, and decision making cannot and should not be ignored. The questions I want to raise in this commentary concern the implications of this research for the rationality debate. Here, I think, there are issues in the wider literature which pose problems for the implications they draw from their work.

To begin with, we should note that there is a need for caution in interpreting one of the core statistical observations that the authors report, the correlation of reasoning task performance with measures of general intelligence. The prime measure referred to is SAT score, but it turns out (n. 3) that the SAT contains a battery of tests, many of which involve reasoning of some kind. It is perhaps then not surprising that correlations are obtained between one block of reasoning tests and another; a truly independent test of general intelligence does not appear to have been used. However, for the time being it is useful to take the claim of a correlation with general intelligence at face value. Happily, there is another overall statistical pattern available: the intercorrelation between (some) reasoning tasks – but it is the ones that do not fit this pattern that are equally interesting.

Let us focus on tests of "practical" reasoning, loosely defined, since there is a well-known aspect of this form of thought which comes under the category of performance that does not correlate either with other reasoning tasks or with SAT scores. This is deontic reasoning, as tested by Wason's selection task. Since at least the seminal work of Cheng and Holyoak (1985), it has been claimed that people display a greater facility with the deontic task than with the indicative or nondeontic form: the majority of experimental subjects seem to produce a normatively justified response to the deontic form, while the majority do not do so with the indicative form (though what is normative in the latter is controversial, in the light of the information gain approach; see Oaksford & Chater 1994).

That people do "better" on deontic tasks, and that this performance is weakly related, if at all, to performance on other tasks and to general intelligence, raises several issues to do with how this performance is characterized. S&W attribute subjects' success to the fact that "in deontic problems, both deontic and rule-based logics are cuing construals of the problem that dictate the same response." Aside from asking how logics can be said to cue anything in an experiment, there is an objection to this idea in the shape of the perspective effect. If we take a deontic conditional of the form "If *p* then may *q*" (a conditional permission), it is possible to specify a range of violations, apart from the one which seems to coincide with the falsifying case for a material conditional, "If *p* then *q*." In the latter instance, that case is *p* and not-*q*. For a deontic conditional, a violation can also consist of not-*p* and *q*, the mirror image of the falsifying case. The difference depends on perspective. For instance, a mother says to her son, "If you tidy your room then you may go out to play": the mother cheats by not letting the child out even though he has tidied his room (*p*, not-*q*) while the son cheats by going out to play without having tidied his room (not *p*, *q*) (cf. Gigerenzer & Hug 1992; Politzer & Nguyen-Xuan 1992). Violation of a deontic conditional is not equivalent to falsification of an indicative conditional, so different norms must apply: there is nothing remotely irrational about the not-*p*, *q* response from the second perspective.

In previous work (see Manktelow & Over 1995; Over & Manktelow 1993) we have argued that deontic reasoning can be usefully considered in terms of decision making: people are sensitive to violations because of the preferences they must mutually assume in deontic discourse. The mother prefers a tidy room, the child prefers to go out. Thus, deontic thought is goal-directed: people are thinking about maximizing their expected utility. Far from settling the question of what is rational is this context, however, invoking decision making opens up new ones. This is because it may be impossible, in principle, to state what is rational in decision making, because a given decision can further one goal at the expense of another: goals can conflict, leading to intractable dilemmas.

S&W state that "construals consistent with normative rationality" – which they link to higher cognitive ability – "are more likely to satisfy our current individual goals." These two words – current, individual – are the crux. Sometimes we may prefer to satisfy long-term goals rather than seek immediate benefit, as when we take out a pension plan rather than spend the money today, or refuse that extra beer to avert tomorrow's hangover. Sometimes we may prefer to satisfy societal goals at the expense of our own, as when we use public transport rather than a car, or buy more costly "green" products. A testable guess is that people of higher cognitive ability might have a greater tendency to favor the long-term and the societal interest, rather than the immediate and individual. If so, why? And on what grounds would we say they were more rational in doing so? The point is not so much that S&W's System 2 predicts the opposite tendency, but that it is difficult, perhaps impossible, to state in any formal sense what is the normative or rational choice here. What algorithm tells us to favor the long-term over the short-term, or the collective over the individual, or vice-versa? We must appeal, it seems, to our personal values, and this leads us to a trap-door marked "relativism."

Differences, games, and pluralism

Roger A. McCain
Department of Economics and International Business, Drexel University, Philadelphia, PA 19104 **mccainra@drexel.edu**
william-king.www.drexel.edu/WKHP.html

Abstract: While concurring that the evidence from individual differences supports a Meliorist view in the rationality debate, this commentary suggests that (1) the evidence does not clearly support a two-systems interpretation against an interpretation in terms of idiosyncratic differences in mental models, and that (2), especially where interactional processing is concerned, evidence from experimental game theory should also be considered.

In drawing inferences from individual differences in experimental studies of rationality, Stanovich & West (S&W) have contributed and demonstrated a major new tool for the investigation of rationality. In such a sweeping and important contribution, there will be many points that need further discussion. This commentary will be limited to the intersection of two critical points: (1) a potentially relevant body of literature, experimental game theory, has been overlooked, and (2) the dual process theory that the authors support may itself be an instance of "the tendency to see . . . pattern in situations that are . . . unpatterned" (sect. 6.3, para. 1).

The alternative to the two-systems approach is at once unitary and pluralist. It sees a common, evolved human tendency to construct highly idiosyncratic and plural "subjective knowledge structures" (Boulding 1956), some of which deal more effectively with certain experimental problems than others. S&W seem to dismiss this as an instance of performance errors and thus contrary to the evidence that errors are predictable. However, inappropriate mental models may lead to predictable errors, as a game-theoretic example to follow will illustrate.

Game-theoretic models introduce a dimension not usually found in cognitive psychology, "because the environment in which each individual gains experience includes the other players, whose behavior changes as they, too, gain experience." (Erev & Roth 1998, p. 849). If system 1 is indeed social and interactional, experiments based on game examples should clarify its relation to System 2 (McCabe et al. 1996; Fischbacher & Falk 1999).

Consider the following games:

Game A

		Player 2	
		cooperate	defect
Player 1	cooperate	3,3	1,1
	defect	1,1	2,2

Game B

		Player 2	
		cooperate	defect
Player 1	cooperate	3,3	1,4
	defect	4,1	2,2

Game B is the Prisoner's Dilemma, probably the most familiar example in game theory. Game A, however, is different. It is a coordination game with Nash equilibria in case both parties choose "cooperate" or both parties choose "defect." Thus, each player must conjecture the other player's strategy in order to choose his own best response, a difficulty that parallels the problem of rational task construal. In this case, however, there are good grounds for a rational conjecture. The fact that "cooperate, cooperate" is better for both players than the other equilibrium makes it a "focal point" (Schelling 1960); each can assume that the other will notice its special character and choose "cooperate," so that the best response to this behavior is also to choose "cooperate."

A mental model to capture this reasoning can be expressed as a "rationale" (McCain 1992):

R:
(I) I want the highest payoff,
(II) Choosing "cooperate" gives both me and my counterpart the highest payoff,
(IIa) and my counterpart, knowing that, will choose "cooperate"; so that (IIb) my own maximum payoff comes from a choice of "cooperate."
(III) Choose "cooperate."

A player with a rich repertoire of game-theoretic mental models will recognize that (IIb) does not apply to game B, the Prisoner's Dilemma game, and will apply a different rationale to it. However, a player with a less rich set of mental models for game theory may fail to make the distinction between the two games and apply to both the simplified rationale:

R*:
(I) I want the highest payoff,
(II) Choosing "cooperate" gives both me and my counterpart the highest payoff, so that is the best choice.
(III) Choose "cooperate."

This oversimple mental model still leads to the best (normatively rational) outcome in the coordination game, but it leads to frequent choices of "cooperate" in the Prisoner's Dilemma game. If the player is lucky enough to be matched with another player who makes the same mistake, they will both be better off than they would if they were rational, which is the fact that gives the Prisoner's Dilemma its fascination. As a result, these mistakes may be mistaken for altruistic behavior or for a higher rationality. But the point for our purpose is that they are predictable: we expect a much greater frequency of "mistakes" on the Prisoner's Dilemma than on the coordination game.

In general, failure to make appropriate distinctions in the mental model will lead to inappropriate contextualization, one of the characteristics of "System 1." This will be correlated with a smaller and less expert set of mental models for game interactions, which is likely to be correlated with many other things.

In conclusion, it seems correct that examination of individual differences will support a Meliorist as against a Panglossian or Apologist position on rationality. However, these observations are consistent with a pluralist view that sees non-normative responses as arising from idiosyncratic subjective knowledge structures. Such a view should not be lumped with performance errors since it is not at all clear that errors stemming from idiosyncratic knowledge structures will be random. Instead, we should investigate what range of predictable errors might arise from idiosyncratic knowledge structures. In doing so, we may draw on a wide range of evidence, including experimental game theory.

Diversity in reasoning and rationality: Metacognitive and developmental considerations

David Moshman

Department of Educational Psychology, University of Nebraska, Lincoln, NE 68588-0345. **dmoshman1@unl.edu**

Abstract: Tasks in the adult reasoning literature are designed so that heuristic processing leads one astray and adequate rule-based processing requires explicit knowledge about applicable logical and quasi-logical norms. Other research, however, indicates that appropriate rule-based inferences can be automatic. Individual differences in rationality are largely due to differences in developmental progress toward metacognitive understanding of both heuristic and rule-based inferences.

Theorists of human reasoning have typically assumed that there exists a prototypical way people think and that the goal of psychological research on reasoning is to determine what that way is. Although evidence for diversity in reasoning has long been abundant, it has typically been dismissed as artifactual or theoretically uninteresting. In an important and convincing challenge to the standard view, Stanovich & West (S&W) have demonstrated that, on the contrary, diversity in reasoning is genuine, substantial, systematic, and theoretically important. In this commentary, I elaborate on the nature and locus of diversity in reasoning.

Central to S&W's analysis is a distinction between automatic heuristic processing (characteristic of what they call System 1) and explicit rule-based processing (characteristic of what they call System 2). I believe this dichotomy confounds two orthogonal distinctions. Specifically, the distinction between automatic and explicit processing is conceptually orthogonal to the distinction between heuristic and rule-based processing. Crossing automatic versus explicit with heuristic versus rule-based suggests four possible types of processing: (a) automatic heuristic processing (System 1), (b) automatic rule-based processing (not represented in the Stanovich/West analysis), (c) explicit heuristic processing (also not represented), and (d) explicit rule-based processing (System 2).

Why do S&W collapse the two distinctions into one, and thus end up with two categories rather than four? I think it is because they focus on the literature on adult reasoning. On the tasks presented to subjects in this literature, heuristic processing tends to be automatic, whereas rule-based processing requires explicit awareness and control of one's inferences.

Research on elementary logical and mathematical inferences, however, shows that people of all ages, including preschool children, routinely make automatic inferences that are fully in accord with rules of deductive logic, probability theory, and so on (Braine & O'Brien 1998; Hawkins et al. 1984; Huber & Huber 1987; Scholnick & Wing 1995). Without a steady stream of unconscious rule-based inferences, in fact, ordinary activities such as reading and conversation would be impossible.

Correspondingly, research and theory on metacognition suggest that explicit reasoning often involves the deliberate application of heuristic principles (for reviews, see Kuhn 2000; Moshman 1998; 1999). In fact, if I may momentarily construe Stanovich & West as research subjects, the arguments they provide in their target article (and similar analyses by authors they cite) constitute clear evidence that human beings are capable of reasoning on the basis of explicit understanding about the advantages and limitations of various heuristic strategies.

Putting all this together suggests that, beginning in the preschool years, all individuals routinely make a variety of automatic inferences, both heuristic and rule-based. Over the course of development, to varying degrees, people increasingly engage in explicit reasoning. That is, they increasingly deploy and coordinate heuristic and rule-based inferences on the basis of increasing metacognitive knowledge about the nature, applicability, and justifiability of various forms of heuristic and rule-based inference (Kuhn 2000; Moshman 1994; 1998; 1999). This picture has several important implications for our understanding of human rationality that are consistent with S&W's emphasis on diversity but go beyond their focus on individual differences.

First, without denying the importance of differences across individuals, it appears that a great deal of the diversity in human reasoning exists *within* individuals. From early childhood, people routinely process information, automatically and unconsciously, in accord with a variety of norms. Some of these norms are heuristic guidelines and some are strict logical or mathematical rules. Perhaps some people are more disposed toward heuristic processing and some toward rule-based processing but all people at all ages regularly engage in both. With regard to the distinction between heuristic and rule-based processing, the primary locus of diversity is within individuals.

Second, differences *across* individuals appear to be largely developmental. Over the course of childhood, adolescence, and early adulthood, people increasingly – but to differing degrees – recognize that some inferences are better than others and that their conclusions and actions will be more justifiable if they constrain their inferences in accord with appropriate norms. Thus, they construct increasingly explicit knowledge about the nature and applicability of various heuristic and rule-based norms and, on the basis of this knowledge, are increasingly deliberate in their reasoning. Although automatic inferences are ubiquitous across the lifespan, there is a developmental trend toward increasingly explicit reasoning.

Finally, the present developmental picture suggests that rationality is fundamentally a matter of metacognition and only secondarily a matter of conformity to various logical or other norms. Individuals who deliberately choose to apply a particular rule, principle, framework, or metaphor on the basis of an explicit understanding of the advantages and limitations of various normative and strategic options are functioning as rational agents, even if they make mistakes in the course of their deliberations. Their rationality can be evaluated, in fact, precisely because it possible for them to make mistakes. As metacognitive agents, they can be held responsible for their inferences.

In contrast, a computer that automatically processes information in accord with its program is not a rational agent at all, even if its processing of information is fully in accord with logical or other rules (Moshman 1994). Its rationality cannot be meaningfully evaluated. If it were to generate unjustifiable conclusions, responsibility for the faulty processing would lie with the programmer, not with the computer. The question of rationality arises only with regard to agents who are sufficiently metacognitive to make deliberate inferences and thus to be responsible for their processing of information.

In summary, Stanovich & West have provided a valuable picture of individual differences in rationality. Extending this picture will, I think, require greater attention to diversity *within* individuals, the metacognitive nature of rationality, and the developmental basis for individual differences in metacognition.

Are there two different types of thinking?

Stephen E. Newstead

Department of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom. **s.newstead@plym.ac.uk**

Abstract: Stanovich & West's claim that there are two coherent and conceptually distinct types of thinking, System 1 and System 2, is questioned. Some authors equate System 2 with intelligence whereas other do not; and some authors regard the two types of system as distinct while others regard them as lying on a continuum.

There can be no question that Stanovich & West (S&W) have made an important contribution to research on reasoning by emphasising the existence of individual differences and the implications they have for theories of reasoning. In this commentary I wish to focus on just one issue: the claim that there is in the literature a wealth of evidence supporting the assertion that there exist two quite distinct types of thinking.

In their Table 3, S&W list a variety of authors who have postulated the existence of two systems. These include the distinctions between heuristic and analytic processing, implicit and explicit learning, and experiential and rational thinking. S&W make the following claim: "Although the details and technical properties of these dual-process theories do not always match exactly, nevertheless there are clear family resemblances." In the ensuing discussion they treat the two types of thinking process, which they label System 1 and System 2, as though they are two quite distinct and conceptually coherent systems. Although they are not the first authors to make this claim (see, for example, Epstein et al. 1996), and although it is true that there are striking similarities between the distinctions made by different theorists, I wish to claim that there is little or no evidence that they amount to the same thing, and considerable reason for believing that they do not.

In their target article, S&W present no argument or evidence that the same distinction is being made by all these authors. Presumably, they thought that the similarities were so transparent not to require such a defence. It is appropriate, then, to ask what kind of evidence would support their claim. Perhaps the most persuasive line of evidence would be the existence of high correlations between all these different measures. Unfortunately, few such studies seem to have been done. Several studies have looked at correlations between subcomponents of the two, but with inconclusive results. For example, Epstein et al. (1996) found that superstitious and categorical thinking, which might be supposed to be part of System 1, produced no significant correlations, either positive or negative, with Faith in Intuition (System 1) or Need for Cognition (System 2). Stanovich & West (1997) themselves looked at correlations between various measures of thinking which might be related to either System 1 or System 2, but reported only "moderate intercorrelations" (their phrasing). In any case, there are conceptual problems here since it is far from clear just how high a correlation would be needed to provide evidence that two types of thinking are part of the same System.

A more revealing line of evidence derives from systematic correlations between System 1 versus System 2 thinking and other psychometric measures such as general intelligence (*g*). Many of the researchers included in S&W's Table 3 would actually equate System 2 thinking with general intelligence. Evans (2000) states quite explicitly: "Rationality$_2$ involves individual differences in *g* . . . Hence intelligence – in the sense of *g* – depends upon the effective use of the explicit thinking system."

Others, however, set considerable store by the claim that their version of System 2 thinking is *not* the same as intelligence. Klaczynski and his colleagues have carried out a series of studies investigating relationships between individual differences in rational processing and intelligence, and performance on a variety of reasoning tasks. In the light of their finding that there are few correlations between measures of rational processing and intelligence, Klaczynski et al. (1997) drew the conclusion that "decontextualized reasoning is a function of an array of personal dispositions distinct from intelligence" (p. 481). We have confirmed in our own laboratory (Handley et al., 2000) the claim of Epstein and his colleagues (e.g., Pacini & Epstein 1999) that their measure of rational thought does not correlate with standard intelligence scores. The fact that one supposed type of System 2 thinking is the same as intelligence while another is completely distinct from it surely leads to the conclusion that they are not really part of the same system.

Other aspects of the different types of thinking presented in Table 3 also lead to suspicions that they are really quite different things. Some of the distinctions represent true dichotomies, in the sense that they are mutually exclusive categories. System 1 thinking is unconscious (tacit, implicit) while System 2 thinking is conscious (explicit). This does not seem to permit any half-way house, that is, thinking that is partly conscious. However, other types of System 2 thinking would appear to lie on a continuum with System 1 thinking. There is a continuum between automatic and controlled processing; indeed, one of the most widely used examples is that of driving a car, where a skill which is at first highly controlled gradually becomes more automatic. Similarly, the distinction between fast and slow processing is a difference of degree rather than kind.

A related difference between the various types of thinking lumped together in Table 3 involves the independence of the processes. As we have seen, some of the dimensions are clearly regarded as being related (for example, automatic and controlled processing are usually considered opposite ends of a single dimension). However, Epstein and his colleagues (e.g., Pacini & Epstein 1999) argue that their dimensions of rationality and experientiality are unipolar rather than bipolar. Once again, our own research (Handley et al. 2000) confirms that these dimensions are indeed independent of each other.

It might be assumed from my argument that I do not believe in the existence of two distinct types of thinking. My view is more one of agnosticism than of disbelief. I do not believe that there is good evidence, other than simple intuition, that the types of thinking regarded as similar to each other really are conceptually related. It is an empirical matter whether or not this is the case, and studies looking at the relationships between the different measures are clearly required. I would not be surprised if some of the measures correlated well with each other, but I would be surprised if they all did. Pending the publication of such research, Table 3 must remain little more than a fiction – a convenient and attractive one, but a fiction nevertheless.

ACKNOWLEDGMENT
The author was in receipt of research support from the Economic and Social Research Council, United Kingdom, at the time of writing this commentary.

Paradoxical individual differences in conditional inference

Mike Oaksford and Jo Sellen
School of Psychology, Cardiff University, Cardiff, CF1 3YG, Wales, United Kingdom **{oaksford; sellenjl}@cardiff.ac.uk** **www.cardiff.ac.uk/psych/Oaksford**

Abstract: Paradoxical individual differences, where a dysfunctional trait correlates positively with some preconceived notion of the normatively correct answer, provide compelling evidence that the wrong norm has been adopted. We have found that logical performance on conditional inference is positively correlated with schizotypy. Following Stanovich & West's reasoning, we conclude that logic is not normative in conditional inference, the prototypically logical task.

In assessing the implications of individual differences in reasoning for the rationality debate, Stanovich & West (S&W) concentrate on differences in intelligence. However, there are other individual differences that are associated with differential reasoning performance. Here we consider one example and show that following the consequences of S&W's arguments leads to an interesting conclusion concerning the normative status of logic for conditional inference.

Conditional inference is the task that most directly tests people's logical competence using the most important construct in logic, the *if . . . then* statement. In this task, participants assess inferences from conditional rules, *if p then q.* Two inferences are logically valid: modus ponens (MP), *if p then q, p,* therefore, *q;* and modus tollens (MT), *if p then q, not-q,* therefore, *not-p.* Performance is logical if participants endorse MP and MT and refrain from endorsing fallacies. Experimental participants tend to endorse MT significantly less often than MP and they also endorse fallacious inferences.

Accounting for this behaviour focuses on the fact that we normally have other information available in semantic memory that influences our inferences (e.g., Byrne 1989; Cummins et al. 1991). For example, take the rule, *if you turn the key the car starts.* If I believe that the car won't start because it is out of fuel, or the ignition is broken, and so on, then I might be less willing to conclude that *the car starts* on learning that *the key has been turned* (MP), or that the *key has not been turned* on learning that *the car has not started* (MT). That is, people should be less willing to make these logical inferences when exceptions are available; this is what both Byrne (1989) and Cummins et al. (1991) found. This result is not specific to the causal relations used in our example but occurs across all domains that have been used in conditional inference (Thompson 1994; Thompson & Mann 1995).

This pattern of results has suggested a direct connection between reasoning ability and schizophrenia. Schizophrenic patients are prone to "overinclusive thinking" (Cameron 1939; 1954; Chapman & Chapman 1973), for example, they may categorise "aeroplane" as an instance of the category "bird" (Chen et al. 1994). This deficit is consistent with ignoring exceptions: these patients obey the general rule "if it flies it's a bird" but ignore the exceptions, that is, unless it's a plane, or a bat, and so on. If nonlogical performance on the conditional inference task occurs because of the availability of exceptions, then schizophrenic patients, who ignore exceptions, may show more logical performance. Schizophrenia seems to be one end of a continuum of problems that occur in the general population. We therefore tested this hypothesis by using schizotypy scales to assess levels of this trait in a normal sample. The high schizotypy group made significantly more MP and MT inferences than the low schizotypy group. Moreover, overall, schizotypy was strongly positively correlated with logical performance. That is, paradoxically, logical performance was positively correlated with high levels of what is usually regarded as a dysfunctional psychological trait.

How are we to interpret these results? S&W provide some interesting answers. As they mention (sect. 4.3, para. 5), Funder (1987) argued that certain attributional errors were associated with better social adjustment and that consequently "the so-called error is thus probably produced by a judgmental process that is generally efficacious" (sect. 4.3, para. 5). As S&W point out, these results can be interpreted as indicating that the wrong normative model has been applied to the task. This interpretation suggests that we should regard the positive relationship we observed between schizotypy and logical reasoning in the same way: logic is not the appropriate normative model of conditional inference. This interpretation is consistent with Oaksford et al.'s (in press) recent probabilistic account of the conditional inference task. They show that performance appears rational when compared to probability theory, which naturally incorporates exception information.

However, S&W point out that performance on many reasoning tasks is positively correlated with intelligence (sect. 4.4, para. 1). Such findings seem to imply that the standard normative model is normative for many tasks because more computation power is unlikely to be associated with non-normative responding. Consequently, if schizotypy had been positively correlated with intelligence, then these results show without question that logic is the appropriate normative model of conditional inference. We did not assess the IQs of the participants in this study, however, there are good reasons to doubt that such a relationship would be observed. First, it runs counter to the theory that suggested the possibility of a relationship between conditional reasoning performance and schizotypy. According to that theory, an inability to access exception information would lead to logical performance. However, it seems reasonable to assume that increased computational power will allow more access to relevant information not less. Second, it seems highly unlikely that schizotypy correlates positively with intelligence and increased computational power. The schizotypy scale that correlated most highly with logical performance was the impulsive-nonconformist dimension. Impulsivity is normally associated with allocating insufficient cognitive resources to a task, hence with reduced rather than increased computational power. Third, our findings were consistent with other studies showing that schizotypy is associated with attentional deficits and lower IQ scores (e.g., Obiols et al. 1999).

Following S&W's reasoning, this example illustrates that the study of individual differences in reasoning has the potential to provide new insights into an often confusing set of findings. It also illustrates that it is interesting to study a broader range of individual differences than just intelligence. Until now, most studies of individual differences in mainstream reasoning research have concentrated on differences in reasoning *strategy* (e.g., Bucciarelli & Johnson-Laird 1999; Ford 1995). However, many of these performance differences may relate directly to personality and other individual difference measures that we have only just begun to investigate. Moreover, 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. How could increased ability to produce the rational response result from increased levels of psychological dysfunction? The only conclusion appears to be that the wrong norm has been applied and that the behaviour that is normative is to be found at the other end of the scale measuring the dysfunctional trait.

Do we need two systems for reasoning?

Klaus Oberauer
Allgemeine Psychologie I, University of Potsdam, 14415 Potsdam, Germany
ko@rz.uni-potsdam.de
www.psych.uni-potsdam.de/staff/oberauer/index-e.html

Abstract: The hypothesis of two separate reasoning systems, one subserving individual goals and the other our genes, is theoretically implausible and not supported by the data. As an alternative, I propose a single system for analytical reasoning backed up by simple mechanisms for the selection of relevant information. This system can generate normative behavior as well as systematic deviations from it.

Stanovich & West (S&W) propose a very strong version of a dual-process account of reasoning. They not only distinguish two kinds of processes characterized by different values on a number of property dimensions, but they suggest that these processes arise from two separate systems of reasoning, each of which produces its own answers to a problem. Moreover, these systems are assumed to serve different goals – those of the individual and those of the genes, respectively.

I think the evidence does not support such strong hypotheses. The positive manifold among reasoning tasks does suggest that the processes that lead to a normative answer have something in common, consistent with the idea of a general system for analytical reasoning. This does not imply, however, that the systematic deviations from normative responses in different reasoning tasks also arise from a single system – they might have no more in common than the absence of sound analytical reasoning.

S&W (sect. 6.1) base their position on results from four problems, the Wason selection task, the 2 × 2 contingency assessment problem, the Disease Problem (associated with the framing effect), and the Linda problem (associated with conjunction fallacy). With respect to the first two, adaptionist accounts have correctly pointed out that it is rational, given certain general assumptions about the environment, largely to ignore negative instances (i.e., not-P and not-Q cards, events with cause and effect absent). This is no more rational, however, from the point of view of our genes than from our individual point of view. Moreover, when looked at in more detail, the reasoning process does not show the signs of sophisticated adaptation that would be expected from a separate system for evolutionary rationality. For example, although people are occasionally sensitive to the relative frequencies of P and Q cases in the Wason selection task, they do not at all show the precise patterns of effects predicted by Oaksford and Chater's rational analysis (Oaksford et al. 1999; Oberauer et al. 1999). The other two phenomena (the framing effect and conjunction fallacy) are linked only indirectly to "evolutionary rationality" through the assumption that conversational implicatures serve evolutionary goals. This is certainly true, but only in the trivial sense that following pragmatic principles of communication helps in communicating successfully, which in turn serves individuals as well as their genes. It is hard to see how following conversational implicatures serves one's genes, while not following them serves one's individual goals.

I believe that the normative responses to reasoning tasks as well as the systematic deviations from them can be explained within a single reasoning system. Every efficient analytical reasoning system must rely on some mechanism to select its input. No task is completely decontextualized; even if the instructions state explicitly that only the information given should be used, a task must be linked to existing knowledge. Without retrieval of relevant knowledge the person would not even understand the information given. In the Disease Problem, for example, participants are expected to draw on their knowledge that scientific estimates are valid information and that people cannot both die and be saved; they should even infer from "200 people will be saved" that the other 400 will die (note that in this case, following a conversational implicature is crucial to reaching the analytical answer!). On the other hand, people are expected not to use their knowledge that scientific estimates are never that exact. So, to understand the task, a lot of background knowledge is necessary, and at the same time it is necessary to exclude other background knowledge from the reasoning process. In addition, reasoners must construct an abstract representation of the information given, which again involves separating the relevant from the irrelevant, and they must select among various formal systems and rules the one adequate to compute an answer. Analytical reasoning is based on a number of implicit decisions about relevance.

If decisions about relevance in a large knowledge base were to be done in an analytical, rule-based way, the system would be doomed to run into the frame problem. Relevance decisions will most likely be based on a quite primitive but efficient mechanism, for example, spread of activation in a semantic network. This idea is implemented in ACT-R (Anderson & Lebiere 1998), where the momentary activation of each knowledge chunk represents an estimate of the likelihood that the element will be used in the context of the information that is presently in the focus of attention. Such a mechanism is, of course, evolutionarily adaptive. But it does not serve our genes as opposed to our individual goals, and it is not a separate reasoning system that provides answers to problems on its own. Instead, the gradient of activation over knowledge units provides the basis on which an analytical system can reason.

A mechanism that selects relevant information can fail, and it will fail in a quite systematic way when the cues for relevance that are usually valid are invalid in a particular case. The four problems discussed above can be understood in this way. A spreading activation system such as ACT-R will give the concepts mentioned explicitly in the problem description strong activation. This leads naturally to a "positivity bias" (Evans 1989). We might say that the system works with the "assumption" that positive instances of categories are more important than negative instances – which is usually correct. The assignment of higher relevance to positive instances explains the selection of the positive cases P and Q in the Wason task and the higher weight of positive compared to negative evidence in the contingency table task. A quite similar positivity bias seems to be at work in the Disease Problem: A description involving "200 people will be saved" will strongly activate the concept "save," whereas the description "400 people will die" directly activates the concept of death. The negations of these concepts, respectively, will not receive the same activation unless the system performs an extra inferential step from 200 people dying to the survival of the rest, and vice versa. The same simple principle also leads to the erroneous assumption that the elaborate description of Linda is relevant, which is at least one factor in the conjunction fallacy (Hertwig & Gigerenzer 1999). More specifically, the description of Linda will activate the concept "feminist" more than the concept "bankteller." If the reasoning system takes the activation values of concepts in the two propositions directly as hints about which proposition to select, this will generate the conjunction error.

Thus, the very system that supports rational reasoning by selectively activating relevant input for it will introduce systematic biases when it is manipulated by varying the simple cues it uses. When the reasoning system is powerful enough (i.e., when it has sufficient working memory capacity to inhibit irrelevant information or compute additional relevant information, when it has access to the right formal rules, etc.), it can overcome these biases. When it is not, it will arrive at decisions that would be correct if the most highly activated information were indeed the most relevant one.

What appears as "radical contextualization" might just be the common denominator of various failures to distinguish relevant facts and rules from irrelevant ones. The irrelevant information or procedures brought to bear on a problem will, of course, always be of the kind that would be relevant in slightly different circumstances and, hence, can be interpreted as "rational" or at least adaptive with a little bit of goodwill.

When we elaborate the system for analytic reasoning in the way sketched here, we can see that its failure will not result in random errors. Instead, we should expect quite systematic non-normative responses from the intrusion of irrelevant information, together with some dumb fallback rules like "choose the case that is most highly activated." This should be the baseline model against which to test the assumption that a second reasoning system is at work to generate non-normative responses.

Bayes, Levi, and the taxicabs

Samir Okasha
Department of Philosophy, Logic and Scientific Method, London School of Economics, London WC2A 2AE, England **s.okasha@lse.ac.uk**

Abstract: Stanovich & West (S&W) are wrong to think that all "reject-the-norm" theorists simply wish to reduce the normative/descriptive gap. They have misunderstood Issac Levi's reasons for rejecting Tversky and Kahneman's normative assumptions in the "base-rate" experiments. In their discussion of the taxicab experiment, (S&W) erroneously claim that subjects' responses indicate whether they have reasoned in accordance with Bayesian principles or not.

Stanovich & West (S&W) rightly emphasize the need for principled constraints on alternative explanations of the normative-descriptive gap. I am sympathetic to much of what they say concerning the relevance of patterns of individual differences, but shall concentrate on two points of disagreement.

"Reject-the-norm" theorists are those who argue that subjects' responses have been judged against incorrect normative models of reasoning/decision-making. S&W claim that all "reject-the-norm" theorists are Panglossians, who oppose particular normative models *because* actual performance fails to accord with them (sect. 4.1, para. 1). This is incorrect, however. Some authors reject the norm for independent reasons. Issac Levi's influential (1983) paper, from which S&W actually quote, is a case in point. Levi argues that Tversky and Kahneman (1977) use a normatively incorrect model in their "base-rate fallacy" experiments. Levi shows that, given their view about which response is normative in three different versions of the taxicab problem, Tversky and Kahneman must be imposing constraints on prior probability allocations which are inconsistent with the Bayesian principles they claim to espouse. Levi is *not* opposing Tversky and Kahneman's normative account on the grounds that it diverges from actual performance, as S&W appear to think. Indeed, Levi himself suggests that Tversky and Kahneman's experimental subjects probably *are* reasoning badly (p. 505). Levi rejects Tversky and Kahneman's normative account because of its internal inconsistency, not because of an *a priori* belief that the normative and descriptive must co-incide. S&W are quite wrong to classify Levi as a Panglossian; reject-the-norm theorists come in different stripes.

S&W's failure to see that not all reject-the-norm theorists are Panglossians has implications for their argument in section 4.4. Invoking the understanding/acceptance principle, they claim to "embarrass" reject-the-norm theorists by pointing to the correlation in Table 1 between higher intelligence and the tendency to give the response generally considered normative. Surely we would not want to argue that subjects who are higher in cognitive ability systematically compute the non-normative response, they state rhetorically. But this correlation is only embarrassing for those reject-the-norm theorists whose sole motivation is to rescue subjects from the charge of irrationality, that is, Panglossians. (Indeed, for such theorists, the fact that individual responses vary systematically in the first place, let alone that the variance correlates with cognitive ability, is embarrassing enough – for it immediately complicates the attempt to read off the normative from the descriptive.) For those who reject the norm for reasons that have nothing to do with trying to reduce the normative/descriptive gap, whether or not more cognitively able subjects tend towards a particular response is not to the point.

In their discussion of the taxicab problem, S&W fall prey to a confusion which Levi (1981; 1983) and Niiniluoto (1981) cleared up, but is still unfortunately widespread (sect. 4.5, para. 1) After a brief description of the problem, they write: "Bayes's rule yields .41 as the posterior probability of the cab being blue." They then classify subjects' responses as "Bayesian" if they fall between .30 and .70, and "non-Bayesian" otherwise. But this is most misleading. It is not Bayes's rule per se that generates a posterior probability of .41 that the cab is blue. Rather, it is Bayes's rule plus a particular assumption about how to assess the probability that the cab is blue, given that the cab had an accident and 15% of cabs in the city are blue, (the "prior" probability in Bayes's calculation), which yields a posterior of .41. The assumption in question – that that probability is .15 – has nothing at all to do with Bayesianism. There are arguments purporting to show that the probability in question should be .15, and other arguments purporting to show that it should not be, but all such arguments are strictly additional to the core Bayesian principles – that an agent's personal probabilities should satisfy the probability calculus at any one time, and change from one time to another by conditionalization. So it is quite illegitimate to infer that subjects whose posterior for the cab being blue differs markedly from .41 are reasoning in a "non-Bayesian" way. Indeed, as Niiniluoto and Levi both point out, subjects who use the principle of insufficient reason to set their prior probabilities will judge the probability that the cab is blue, given that it had an accident and 15% of cabs in the city are blue, to be .5; applying Bayes's rule to this prior gives a posterior of .8 for the cab being blue, which is the subjects' modal response. Stanovich & West's exploration of the correlation between cognitive ability and different responses to the taxicab problem is interesting, but their classification of such responses as "Bayesian" and "non-Bayesian" is erroneous. The version of the taxicab problem with which they operate does not test whether subjects are reasoning in a Bayesian manner or not.

Rational distinctions and adaptations

D. E. Over[a] and J. St. B. T. Evans[b]
[a]*School of Humanities and Social Sciences, University of Sunderland, Sunderland SR1 3PZ, United Kingdom;* [b]*Department of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom*
david.over@sunderland.ac.uk **J.Evans@plymouth.ac.uk**

Abstract: Stanovich & West (S&W) distinguish between evolutionary rationality and normative rationality, and System 1 and System 2 mental processes. They hold that the main function of System 2 has to do with normative and not evolutionary rationality. We ask how System 2 could then be an adaptation, especially given S&W's own work on individual differences.

In general we strongly welcome the advances that Stanovich & West (S&W) have made by their application of the method of individual differences to the study of rationality and reasoning. We naturally find their dual process theory very congenial to our own (Evans & Over 1996) and note that their Systems 1 and 2 develop the rudimentary notions of type 1 and type 2 processes first discussed by Wason and Evans (1975). We limit our comments on their paper, however, largely to their distinction between what they call evolutionary rationality and normative rationality. It is better, we think, to describe normative rationality as individual rationality. Both types of rationality are broadly instrumental, as S&W realize. Evolutionary rationality can be thought of as serving the metaphorical goal of the genes, that is, reproductive success. Individual rationality serves the goals of the whole individual, which can be many and varied in different individuals and have to do with avoiding reproduction altogether.

In our own theory, we distinguish rationality 1 from rationality 2. The former results from implicit type 1 or System 1 processes, and the latter from explicit rule following in type 2 or System 2 processes. As S&W point out, our distinction is between different mechanisms – tacit heuristics versus reason-based rule following – for achieving individual goals and so individual rationality. We had already identified the need for an explicit System 2 of hypothetical thinking to account for rationality 2, but we now know from S&W's work that the facility for this rationality is related to measures of general intelligence. The question is how to apply these distinctions usefully to the notion of evolutionary rationality.

S&W are certainly right to emphasize the distinction between evolutionary and individual rationality, as serious confusion may result from the failure of psychologists to pay attention to it. For example, the group they call the cognitive ecologists as well as many evolutionary psychologists presuppose that what has evolutionary or ecological rationality (Todd & Gigerenzer, this issue) also has individual rationality. This group does sometimes argue that the heuristics they try to discover were adaptive under primitive conditions, but still tends to presuppose that these heuristics always have individual rationality in the contemporary world. It would be more plausible to argue, on the contrary, that these heuristics can be the source of biases in advanced technological societies. And however rare the bias, demonstrating its presence is good evidence for the existence of the heuristic. The Panglossian perspective of this group may have prevented them from getting better experimental evidence for their heuristics and making a positive contribution to the heuristics and biases literature.

System 1 heuristics can produce damaging biases for individuals, but we agree with S&W that System 1 not only contributed to evolutionary rationality, but still on the whole facilitates individual rationality. The more interesting question concerns System 2 and the rationality 2 which comes with it. S&W say that System 2 acts more to serve individual rationality than evolutionary rationality, but in some tension with this, they add that System 2 evolved by natural selection. However, consider the technical concept of heritability and its relation to adaptation. Heritability is the proportion of phenotypic variance which is due to the genes. It is arguable that an adaptation should generally have low heritability. For example, take the ability to speak a natural language as an adaptation. It has low heritability, for if a human being fails to have a high level of this ability, that is probably the result of nongenetic influence, like being severely neglected as a child. But the vast literature on intelligence implies that having a high IQ is not like that – much more of the variance in IQ is down to the genes. Thus, it could be suggested that System 2, which S&W link to high IQ, is not an adaptation. Indeed there are evolutionary psychologists who deny the existence of System 2 as an adaptation and reject any dual process theory as an account of the natural human mind (Tooby & Cosmides 1992).

We ourselves hold that a basic level of System 2 ability and rationality 2 contributed to reproductive success under primitive conditions, at least by helping with novel problems (Evans & Over 1997; Over & Evans 1997). But it may be that the development of large brains with the capacity for language and explicit rule following coincided with and contributed to the curtailing of the influence of natural selection on our gene pool. That is, once humans became social and communicating animals with organized social structure, the presence of highly intelligent people, with high System 2 ability, boosted the development of culture and technology, without contributing differentially to their own reproductive success, but benefiting everyone more equally in this respect. In other words, highly intelligent people may have themselves, in effect, prevented their level of System 2 ability from becoming more widespread.

Data, development, and dual processes in rationality

Valerie F. Reyna
Informatics and Decision-Making Laboratory, Departments of Surgery and Medicine, University of Arizona, Tucson, AZ 85724-5031
vreyna@u.arizona.edu

Abstract: Although Stanovich & West (S&W) are likely to be criticized for not proposing a process model, results of such a model (fuzzy-trace theory) support many of their conclusions. However, arguments concerning evolution and Gricean intelligence are weak. Finally, developmental data are relevant to rationality, but contradictory results suggest a dual-processes approach that differs from S&W's based on fuzzy-trace theory.

Stanovich & West (S&W) have brought the discussion of rationality back to data, after a period of mostly post hoc speculation. With due reverence for philosophy, they rightly argue that an entire dimension of data has been ignored, namely, individual differences. A likely criticism of this approach is that it is not process oriented. I accordingly begin by noting results predicted by a process-oriented theory (Fuzzy-trace theory) that support many of S&W's conclusions. Next, I discuss weaknesses of arguments concerning evolution and Gricean maxims that are used to characterize intuitive processes.

Last, I discuss an additional stream of research that S&W claim, on analogy with individual differences, is relevant to rationality: developmental data. Contrary to their claim, however, a major theme of such research is that heuristics and biases increase with age, despite parallel improvements in computational competence. This paradox, along with task analyses in adult reasoning, have been pivotal in formulating an alternative to both the heuristics/biases (irrationality) and the adaptive/ecological (rationality) frameworks, fuzzy-trace theory. Fuzzy-trace theory offers a dual-process account of rationality grounded in memory research that differs in important respects from other dual-process approaches, including S&W's.

S&W provide evidence that diverse reasoning tasks are related to one another and to SAT scores. Computational capacity has been investigated in these tasks, but has been ruled out (for literature reviews, see Reyna 1992; 1995). Limitations of correlational analyses, as used by S&W, have also been discussed (Brainerd & Reyna 1992a; 1992b; Brainerd et al. 1999; Reyna & Brainerd 1990). However, cognitive mechanisms have been found that are operative across tasks. In 1991, Reyna extended fuzzy-trace theory's process model of inclusion illusions to syllogistic reasoning, conditional probability judgment, base-rate neglect, class-inclusion errors, and the conjunction fallacy. Reasoning errors were explained by interference from inappropriate gist representations, from irrelevant reasoning principles, and, most important, processing interference from nested classes (Brainerd & Reyna 1990; 1993; Reyna 1995; 1996; Reyna & Brainerd 1993; 1994; 1995; Reyna et al., in press; Wolfe 1995). Dempster (1992) and others have argued that individual differences in susceptibility to interference account for relations across tasks, consistent with S&W's data.

Consistent with S&W's attribution of framing effects to intuitive processing, research has shown that analytical, quantitative processing reduces such effects and categorical, qualitative processing increases them (Reyna & Brainerd 1991; 1995). S&W ascribe intuitive processing to evolutionary adaptation and interactional intelligence that "support[s] a Gricean theory of communication." The main "evidence" for the evolutionary argument is that subjects give intuitive responses. However, the mere existence of a bias or behavior is not evidence that it is adaptive.

As for Gricean intelligence, similar arguments were made in the developmental literature when children failed Piagetian tasks. The denouement in that literature, as in this one, is that effects remained when ambiguities and supposed Gricean implicatures were controlled for (Reyna 1991; Reyna & Brainerd 1991; Tversky & Kahneman 1983). Unfortunately, post hoc speculations about Gricean maxims are accepted at face value. For example, it is claimed that Gricean maxims encourage the inference that 200 "or more" people are saved in the framing disease problem. However, these speculations actually violate Grice's maxim of quantity (Clark & Clark 1979; Grice 1978). Speakers are not supposed to omit crucial information. For example, failing to mention that the "woman" a man is meeting tonight is his wife violates the maxim of quantity (if you knew she was his wife, you should have said so). Thus, the Gricean prediction is that if more than 200 people might be saved, a cooperative speaker would have said so.

Without the justification of evolutionary adaptation or Gricean implicatures, what evidence remains for dual-process approaches? I have argued that developmental data are a crucial source of evidence implicating dual processes (Reyna & Brainerd 1994, 1995; 1998). Developmental theorists have assumed that development unfolds in the direction of increasing rationality, as knowledge and experience increase (Piaget 1953; Werner 1948). However, framing biases increase with age, and processing becomes more intuitive (Reyna 1996; Reyna & Ellis 1994). Other biases also increase with development, such as availability, representativeness, and noncompensatory decision-making (Byrnes 1998; Davidson 1995; Jacobs & Potenza 1991). Finally, college students fail Piagetian concrete operational tasks such as conservation of mass (Winer et al. 1992). For example, they think that they weigh more sitting down than standing up.

Across the same age range in which the use of heuristics and biases is increasing, computational competence as tapped by traditional cognitive developmental tasks (class inclusion, conservation, probability judgment) is also increasing. These contradictions across ages and across tasks (within individuals of the same age) provide a major motivation for fuzzy-trace theory's approach to rationality (Klaczynski & Fauth 1997; Reyna & Brainerd 1995). Such task variability cannot be reduced to either Type I or II measurement error, arguing against both the Meliorists (competence is overestimated) and the Panglossians (competence is underestimated).

In sum, a decade of adult and developmental data leads to a conception of rationality that differs from those discussed by Stanovich & West, but which reconciles seemingly contradictory results. Key points of departure include the following: (1) Reasoners encode multiple gist and verbatim representations, which confers cognitive flexibility. (2) However, reasoning operates at the least precise level of gist that the task allows, increasingly so with development. (3) This fuzzy processing preference explains why reasoning has been found to be independent of computational capacity. (4) Thus, rationality is identified with gist-based reasoning rather than with precision as in computational or logicist theories. Despite predictable pitfalls, reliance on gist across superficially different problems is essential for achieving descriptive invariance – the fundamental criterion of rationality. (5) Finally, a given response can reflect more or less rationality depending on the processing used to generate it. Levels of rationality are predicted based on task features and the developmental status of the reasoner. Therefore, rationality is not an immutable aptitude of individuals, but changes from task to task and from one stage of life to another.

ACKNOWLEDGMENTS
Preparation of this paper was supported in part by grants from the National Science Foundation (SBR9730143), U.S. Department of Commerce (04-60-98039), National Institutes of Health (P50HL61212), and the Academic Medicine and Managed Care Forum (SPS Log 38347).

An elitist naturalistic fallacy and the automatic-controlled continuum

Sandra L. Schneider

Department of Psychology, University of South Florida, Tampa, FL 33620-8200 **Sandra@chuma.cas.usf.edu**

Abstract: Although a focus on individual differences can help resolve issues concerning performance errors and computational complexity, the understanding/acceptance axiom is inadequate for establishing which decision norms are most appropriate. The contribution of experience to automatic and controlled processes suggests difficulties in attributing interactional intelligence to goals of evolutionary rationality and analytic intelligence to goals of instrumental rationality.

Stanovich & West (S&W) have made an important contribution to the rationality debate by focusing on the value of the individual differences approach in identifying markers of performance errors and computational limitations. Nevertheless, there is a danger of overstating how much the individual differences approach can buy. S&W's operationalization of Slovic and Tversky's (1974) understanding/acceptance assumption equates normative behavior with the response of individuals who are of high intelligence (as measured by SAT scores). Their position is hence an elitist version of the naturalistic fallacy wherein normative status is ascribed to what more intelligent people do. Although we might expect that those with superior intellectual skills will often adopt better decisions strategies, this seems an inadequate criterion for rationality. Where is the reference to actual success in the environment? S&W seem to suggest that rationality should be defined by what intelligent people do without checking to see in each instance whether the intelligent people end up any better off.

At best, the understanding/acceptance assumption can be viewed as a necessary but not a sufficient criterion for rational behavior. Approaches that are regularly eschewed by intelligent people are not likely to be adopted as normative. S&W (sect. 4.2) argue the converse, borrowing from Larrick et al. (1993): "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." This view is oversimplified because it suggests (a) that intelligent people will be in agreement on the best course of action in a given situation and (b) that "general regard" for what contributes to effectiveness across environments can stand in for empirical evidence of decision quality across innumerable specific situations.

Rationality requires objective criteria so that it can be distinguished from mere opinions about what is best. If no such criteria are available, the rationality argument is likely to remain hopelessly moot. In cases such as formal logic, for which proofs are available to establish the validity of reasoning, rationality may be easier to define than in cases such as risky choice, for which acceptability of axioms is hotly debated. There are intelligent people on both sides of the fence. Moreover, locally non-normative strategies may be globally functional (e.g., optical illusions can result from responding to the same cues that enable depth perception), adding to the complexity of identifying reasonable criteria.

S&W's argument regarding the limitations of using central tendency as the basis for defining appropriate norms (see sect. 4.1) can readily be mapped to the norm itself in the case of expected utility maximization. The expected utility (or expected value) of a risky prospect is a measure of the central tendency of that prospect. However, the central tendency of a probabilistic event is almost by definition an inaccurate indicator of what one might actually experience if the event transpired (e.g., in a 50/50 gamble to win $0 versus $200, the expected value of $100 does not inform individuals of what they will receive after a single play of the gamble). Nevertheless, this single descriptive fact is routinely embraced as the only element required to support rational choice.

For better or worse, techniques like expected utility maximization can be taught, and those who earn high scores on tests such as the SAT are more likely to be exposed to these techniques. Hence, we run into the confound between education (i.e., learning experiences) and intelligence. The positive manifold (i.e., the convergence of measures of cognitive ability) as assessed by S&W does not reflect general intelligence independent of experience. Moreover, the meaning of the positive manifold is constrained by the given cultural and historical context. Without exposure to rules of logic or Bayes' theorem, for instance, one could hardly expect performance consistent with these methods. As time passes and perspectives shift, advocated strategies often change. One day, expected utility maximization may be replaced, casting doubt on the implications of differences between those who did and did not use the approach in its heyday.

A provocative suggestion of S&W is that intelligence-based performance differences might reflect a split between processes in the service of evolutionary adaptation and those aimed at instrumental rationality in the individual. The notion that more automatic, contextualized, and intuitive processes support the former whereas more controlled, decontextualized, and analytic processes support the latter seems plausible at first. On closer inspection, it gets more complicated. Because automaticity is closely related to repetition and practice (Newell & Rosenbloom 1981; Shiffrin & Schneider 1977), many of the processes that might initially involve considerable controlled or analytic processing eventually become largely automatic.

Most decision researchers, for example, can instantly recognize problems of syllogistic reasoning and, without much conscious effort, determine whether the form is valid or not. The ability to recognize this general type of problem can be viewed as synonymous with the ability to use the situational context (e.g., the standard format of two premises and a conclusion) to identify the rule or heuristic to be applied. Hence, the process associated with reasoning problems is not without context. When more complex problem solving is considered, the importance of context becomes even more apparent (Chi et al. 1988).

At the same time, it is easy to discount the extended developmental processes required to learn the many rules of social communication (e.g., Pinker 1989; Schober & Clark 1989). Although some of this learning may be relatively effortless, much of it is likely to involve controlled processing. Social communication problems have context in the same sense that reasoning problems do; particular cues must be interpreted as signals specifying the appropriateness of a given rule. The fact that we can run them off automatically may say more about the amount of experience we have had with similar situations than it does about the types of processing that originally contributed to the development of the action sequence. Although the search for evolutionary versus instrumental processes is promising, the distinctions will inevitably be more complicated than a dichotomy based on types of intelligence.

Cooperative versus adversarial communication; contextual embedding versus disengagement

Keith Stenning and Padraic Monaghan
Human Communication Research Centre, Edinburgh University, Edinburgh EH8 9LW, United Kingdom **{keith; pmon}@cogsci.ed.ac.uk**

Abstract: Subjects exhibiting logical competence choices, for example, in Wason's selection task, are exhibiting an important skill. We take issue with the idea that this skill is individualistic and must be selected for at some different level than System 1 skills. Our case redraws System 1/2 boundaries, and reconsiders the relationship of competence model to skill.

The classical competence model of logic can be thought of as defining the extreme limit of disengagement from implicit context in reasoning. Stanovich & West (S&W) have presented an eloquent argument that controlled disengagement is an important cognitive skill, even if in many (perhaps most) circumstances complete disengagement is not a rational goal.

S&W conclude that System 2 skills of controlling disengagement of reasoning from context have to be treated as arising from quite different selection pressures than System 1 skills. We believe that this conclusion stems first from a confabulation of two dimensions: cooperative versus adversarial communication; and explicit versus implicit knowledge. Second, the discussion is further confused by assuming a narrow relationship between competence models and performance, that neglects metaknowledge and the skills of deployment of formal models.

On the former issue, there are two dimensions for classifying reasoning performance which have been inadequately distinguished in this literature. The first distinction is between cooperative and adversarial communication skills. The second distinction is between skills of implicit and explicit reasoning: reasoning in context using all available beliefs versus disengaging from implicit assumptions to reason only from explicit ones.

Successful students are highly skilled at what Grice (1975) called "conversation" (cooperative expositional communication) and also at what we will call "altercation" (adversarial communication in contexts where they need *not* make a precise disengagement of general relevant beliefs from what are explicitly stated to be shared assumptions). Both are skills presumably selected for at all levels. Different students bring different combinations of these skills to bear in different social contexts. Classrooms (and laboratories) where there are highly asymmetrical authority relations between teacher/experimenter and student tend to suppress the engagement of altercation skills.

The second skill of disengaging from implicit contextual beliefs and attitudes, to reason only from explicit assumptions, as something that most students learn in late formal education. Indeed, this learning takes up a rather large proportion of late formal education. These skills have to grow out of the implicit skills that precede them. Communication often requires complex integrations of cooperative and adversarial elements. For example, many academic tests require students to make sophisticated judgments about which Gricean implicatures they still need to cooperatively draw in order to understand the problem, and which they must adversarially avoid in showing that their solution fits all possible interpretations.

These skills of controlled disengagement from the implicit are what S&W describe as System 2 skills, and conclude that they can only be explained as being selected for at some different level from the System 1 skills of contextualised communication, which they, and the field, see as Gricean and cooperative. But there is rather strong selection pressure for exercise of System 2 skills which is being obscured by their characterisation here as decontextualised reasoning. (One might also share Rose's [1998] scepticism about the biological sense, and political origin, of this distinction of levels of selection).

S&W identify System 1 with cooperation; System 2 with decontextualised reasoning; and omit adversarial communication altogether. We would identify System 1 with the implicit skills of both cooperative and adversarial communication, and System 2 with the skill of explicit control of disengagement from context.

Where does the need for System 2 skills most obviously manifest itself? The competence models originate in the study of educational methods for teaching adversarial communication in the court, parliament, and temple. The skills are required when discourse must be interpretable out of its original context, for example, where laws are made or enforced, or sacred texts interpreted. Shared belief and attitudes cannot be assumed, so there is a premium on making assumptions explicit. The need for controlled disengagement from implicit assumption in adversarial communication long predates in human evolution the Greek invention of the theories that formalise these skills. Why characterise these skills as the playthings of individualism? They are the skills most required for societal functioning once groups have language and grow large enough to not share experiences directly. They are skills likely to gain social status and accession to the elite, with the usual genetic consequences.

To pick up the second issue that confuses this debate, what role does a classical competence model play in accounts of the processes of System 1 and System 2 reasoning? Thinking of the System 2 skills of contextual disengagement as modelled by running some classical logical *mechanism* is misleading. It is not low level rules of deduction that underly System 2. Much of the knowledge and skill of System 2 is meta-logical. For example, one might better think in terms of the social skill in disengaging adversarial reasoning from features of the social authority structure – learning to debate with teacher without the social ceiling falling in.

Our redrawing of the distinctions can be made to do real work. We sketch two examples related to the four-card task. First, our redrawing clarifies the similarities and contrasts between competence models in the literature. Chater and Oaksford's (1994) case for the Bayesian model as a competence model in the four-card task explains widespread failure in terms of subjects' assimilation of the task of assessing the four cards as instructed, to that of seeking evidence from populations of cards. Their theory is based on failure to disengage just as surely as is the logical competence model.

The cooperative adversarial distinction also puts a different interpretation on the dominant content effect in four card reasoning. Deontic social obligation content engages the implicit skills of "altercation" between subject and imagined law breaker, where indicative material originating from the authoritative teacher and with no imagined intermediary, supresses the students' engagement of their skills of altercation. These System 1 skills are not to be modelled as encapsulated widgets such as cheating detectors, but rather general skills in the social practice of adversarial communication.

The ability is not general, and neither are the conclusions

Robert J. Sternberg

Yale University, Department of Psychology, New Haven, CT 06520-8205
robert.sternberg@yale.edu www.yale.edu/rjsternberg

Abstract: Stanovich & West rely for many of their conclusions on correlations of reasoning tasks with SAT scores. The conclusions they draw are suspect because the SAT is not a particularly good measure of so-called *g*; *g* is not necessarily causal, SAT scores are no arbiter of what is true, and in any case it is not suprising that reasoning tests correlate with reasoning tests.

In an interesting and innovative target article, Stanovich & West (S&W) use patterns of individual differences in order to draw inferences about the rationality debate. They do not use just any old pattern of individual differences, however. They use correlations with assessments that measure, to an approximation, Spearman's (1927) general factor, or *g*. They seem to be rather proud of the use of such measures and extol their virtues throughout the target article. Both of two data tables in the article relate performance on reasoning tasks to scores on the SAT, a reasonable although probably not ideal proxy for scores on a *g* factor. There are three problems with the logic of this article that undermine the conclusions based on such relationships – conclusions which, unfortunately, appear to be most of the conclusions in the article.

The multidimensionality of human intelligence. First, Spearman's *g* is only one aspect of intelligence, a fact that even Spearman and other diehard believers in *g* have acknowledged (e.g., Carroll 1993; Jensen 1998). Curiously, after emphasizing the use of a single type of correlate to draw conclusions, S&W themselves discuss at length and cite evidence in favor of two-factor theories of intellectual functioning, which involve a more rule-based kind of reasoning and a more associatively-based kind of reasoning. They further note that these two kinds of reasoning seem not to correlate with each other.

Another theory for which there is fairly extensive evidence is Sternberg's (1985; 1997a) triarchic theory of human intelligence, which postulates and has amassed evidence for three relatively distinct aspects of human intelligence. Gardner (1983; 1999) has argued for at least eight and possibly more multiple intelligences.

Even if one accepted the two-process model proposed by Stanovich and many others, the implication would be that there should be at least two kinds of dependent measures – themselves largely uncorrelated – that should be used in order to draw conclusions. S&W create an internal contradiction when they use only a single measure of ability as a basis for forming judgments. At minimum, they would need to measure the two kinds of intellectual functioning to which they refer repeatedly and cite as supported by evidence.

g is not necessarily causal. As S&W point out, the so-called *g* factor is not well understood. There are many different interpretations of it, and none has compelling support over all the others. As the authors further point out, the factors underlying performance on reasoning tasks also are not well understood. Essentially, what S&W are therefore doing is using one poorly understood construct, *g*, to explain another poorly understood construct, reasoning. It is just as plausible to argue that reasoning is behind *g*, or that both reasoning and *g* depend on some third factor.

The fact that there is a correlation between reasoning tasks and the SAT is not terribly surprising, because the College Board, creator of the SAT, views the SAT largely as a test of reasoning. Indeed, the kinds of items it contains – inferring meanings of words from context, verbal analogies, inferring meanings of passages, and mathematical reasoning problems – come close to being prototypical of reasoning problems. What S&W are really showing, therefore, is that reasoning tests correlate with reasoning tests. No higher order conclusions derive from the fact that reasoning tests correlate with each other.

SAT scores are no arbiter of what is true. Third, there is a not so tacit assumption underlying the target article that SAT scores have some kind of mystical status as an arbiter of cognitive abilities. They do not. It is precisely for this reason that the College Board now uses only the term SAT, rather than using the "A" in "SAT" for "aptitude."

S&W's view seems to suggest to them that the SATs are some kind of mystical arbiter of who really knows what they are doing – that people with higher SAT scores must know better. But know what? Empirical evidence suggests that measures of so-called general ability predict school success moderately and life success only modestly. Even Herrnstein and Murray (1994), strong believers in the *g* factor, agree that estimates of *g* account for only about 10% of the variation in diverse measures of life success (Wigdor & Garner 1982). Some psychologists believe this figure is inflated (e.g., Sternberg 1997a). Whatever the figure, to characterize people with high SAT scores as those who should set the norm for what is somehow true or right seems to be off target. People with high SAT scores have high levels of certain kinds of cognitive abilities. They have no monopoly on quality of thinking and certainly no monopoly on truth. An extensive literature shows that people can be excellent everyday reasoners and yet perform modestly on measures of *g*, or may perform well on measures of *g* and do poorly on everyday reasoning problems (Sternberg et al. 2000).

Conclusion. The issues posed in the S&W target article are important. To the extent their resolutions rely on correlations of reasoning tests with SAT scores, however, the resolutions are suspect. SAT scores cannot serve as the arbiters Stanovich and West would like them to be.

ACKNOWLEDGMENT

Preparation of this commentary was supported by the Javits Act Program (Grant No. R206R950001) as administered by the Office of Educational Research and Improvement, U.S. Department of Education. Grantees undertaking such projects are encouraged to express freely their professional judgment. This commentary, therefore, does not necessarily represent the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education, and no official endorsement should be inferred.

Requests for reprints should be sent to Robert J. Sternberg, Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520-8205

The rationality debate: Look to ontogeny before phylogeny

Stephanie Stolarz-Fantino and Edmund Fantino

Department of Psychology-0109, University of California, San Diego, La Jolla, CA 92093-0109 **sfantino@psy.ucsd.edu** **efantino@ucsd.edu**

Abstract: Subjects have a rich history of decision making which would be expected to affect reasoning in new tasks. For example, averaging, a strategy that is effectively used in many decisions, may help explain the conjunction fallacy. Before resorting to accounts based on phylogeny, more parsimonious accounts in terms of ontogeny should be explored.

In their target article, Stanovich & West (S&W) categorize investigators as "Meliorists" if they emphasize systematic irrationalities in human cognition and "Panglossians," if they offer alternative explanations for the gap between normative approaches and subjects' actual performance. However, in our view, S&W have overlooked at least one approach that is helpful in understanding subjects' departures from normative standards of reasoning: application (or misapplication) of strategies that have been effective in the subjects' histories.

This approach is similar to "alternative task construal," one of the explanations for non-normative behavior discussed by S&W. However, while the authors state that alternative task construal locates the problem "within the experimenter," the approach described here focuses on determining what strategies subjects are applying and what factors influence the strategies they select.

In work on the conjunction fallacy, one of the arenas addressed by S&W, much work has shown that averaging models and other similar quantitative models can be successful in predicting subjects' judgments of the likelihood of conjunctions (e.g., Abelson et al. 1987; Fantino et al. 1997; Gavanski & Roskos-Ewoldson 1991; Massaro 1994; Stolarz-Fantino et al. 1996; Yates & Carlson 1986; Zizzo et al. 2000). Averaging implies that the conjoint probability is equal to or lies between the component probabilities, whereas the normative multiplicative model implies that the conjoint probability is equal to or lower than the component probabilities. Tversky and Kahneman (1983) note that "an averaging process . . . may be responsible for some conjunction errors, particularly when the constituent probabilities are given in numerical form" (p. 306). Such a possibility has several implications: (1) If the likelihood of a conjunction is treated as the average of the likelihood of its components, then it should be possible to predict the rated likelihood of the former from the rated likelihoods of the latter; (2) if the conjunction consists of components that are judged about equally likely, then incidence of the conjunction fallacy should be less than when the conjunction consists of components of divergent likelihoods (since, in the latter case, the average likelihood falls more clearly above the less likely of the two components); and (3) while the presence of a framing description should increase the occurrence of the conjunction fallacy by encouraging the assignment of divergent likelihoods, the fallacy should sometimes occur even in the absence of a frame when subjects assign divergent likelihoods to the components. Research has supported all three of these expectations. For example, Fantino et al. (1997), using the functional measurement methodology of Anderson (1981), found that subjects' ratings of conjunction were consistent with a weighted-averaging model. And Stolarz-Fantino et al. (1996) found that subjects demonstrated the conjunction fallacy at a substantial rate even in the absence of a framing description.

It is not surprising that subjects make errors by misapplying strategies that work for other tasks. Averaging, in particular, is a strategy used in many other types of decisions (e.g., Anderson 1981). If subjects' histories with particular strategies are responsible for at least some of their errors on logical reasoning tasks, how will manipulating their histories affect their performance? This has been done directly in research in another of the arenas addressed by S&W, base-rate neglect (Case et al. 1999; Goodie & Fantino 1996). These studies show that base-rate neglect may result from preexisting associations. In comparable experiments (Hartl & Fantino 1996), pigeons, unfettered by preexisting associations, behave optimally. When the base-rate task is altered so that preexisting associations are absent, base-rate neglect is eliminated in human subjects as well (Goodie & Fantino 1996).

We appreciate S&W's contribution to the rationality debate. Their comments on the limitations of several alternative approaches are valuable, with the caveat addressed in this commentary. The analysis of correlations among specific reasoning tasks seems a fruitful approach to understanding reasoning; we are less convinced of the utility of correlations between single tasks and general measures such as SAT scores. Examining individual differences may help us understand the variability that often characterizes results of reasoning experiments (eg, Stolarz-Fantino et al. 1996). S&W's suggestion that reasoning fallacies may be understood from an evolutionary perspective is thought-provoking; however, as we have suggested, it may be more productive first to look to ontogeny – subjects' histories of decision making and problem-solving – rather than to phylogeny for an appreciation of how we reason as we do.

ACKNOWLEDGMENT

Preparation of this paper supported by NIMH Grant MH57127 to the University of California, San Diego.

Intuitive versus analytic abilities: The case of words versus numbers

Karl Halvor Teigen

Department of Psychology, University of Tromsø, N-9037 Tromsø, Norway
karlht@psyk.uit.no **www.isv.uit.no/seksjon/psyk/karlht.htm**

Abstract: The distinction between abstract (rule-based) and contextual (intuitive) thinking is illustrated by studies of numeric versus linguistic expressions of probability. Verbal probabilities are believed to reflect intuitions that can be adaptive and occasionally normative (e.g., counteracting conjunction errors). Stanovich & West's interpretation of analytic thinking in terms of ability suggests a complementary ability perspective on intuitive thinking.

Many of the reasoning problems that researchers give their subjects can be "solved" in more than one way: through an analysis of the abstract principles involved, or through leads suggested by the description of the concrete situation. Sometimes these solutions do not match, revealing a conflict between rule-based (decontextualized) and associative (contextualized) thinking. Stanovich & West (S&W) have done an excellent job in showing that people differ in their ability to use the analytic approach. This ability appears to be related to differences in analytic intelligence, as measured by standard psychometric IQ tests. Contextualized thinking, on the other hand, may follow heuristic principles of conversational and holistic inference.

This interpretation of the normative/descriptive gap opens an avenue not fully explored by the authors. If the two types of reasoning are due to different cognitive abilities, we can discuss not only the advantages of high versus low analytic intelligence, but also the functions of more or less well-developed varieties of intuitive thinking. If the first ability can manifest itself in varying degrees, so can the latter. And just as high levels of psychometric intelligence are preferable to low levels, so good intuitive abilities should be preferable to deficiencies in intuitive thinking.

To illustrate with a different, but related pair of concepts: people may vary in their propensity for statistical versus clinical reasoning. There are more or less sophisticated statistical thinkers, and there may be more or less sensitive clinicians. A conflict between these two ways of thinking can go both ways; not only will a dedicated clinician sometimes neglect base rates and violate the conjunction principle, but some scientists with a one-track statistical mind seem unable to grasp why risk communication can go wrong, or how people can entertain a set of probabilities of independent events that add up to more than 100%.

Windschitl and Wells (1996) have argued that numeric probability estimates may capture people's rule-based reasoning abilities whereas verbal formulations of probability to reflect their "gut feelings" or intuitions about the uncertainties involved. The same conclusion can be drawn from studies of the "equiprobability effect" (Teigen 1988): when several alternatives are described as equally likely, the chances of each of them may be described as "good" despite their low numeric probability. So if Tom applies for a job together with two other, equally qualified candidates, his numeric probability will be correctly assessed by most respondents as 1/3, yet the same respondents will prefer to speak about Tom's probability in positive terms like "a good chance," "entirely possible," and "not improbable," rather than with negative, low-probability phrases like "rather improbable," "somewhat doubtful," or "quite uncertain." If the number of candidates is increased to six, Tom's numeric probability is reduced to 1/6, whereas the tendency to prefer positive verbal phrases stays largely unchanged (Teigen 1999). In this example, people appear to reason according to a normative principle (the classical 1/n rule of equally likely outcomes) when they are asked in terms of numbers, but they switch to a completely different mode of thinking when asked to state the probabilities in words. Windschitl and Wells (1998) suggest that people compare the target outcome to the other individual outcomes, rather than to their sum (the Alternative-Outcomes effect). They may also reason according to a causal, dispositional model: when Tom is perceived to have the necessary qualifications for being hired, and nobody is clearly ahead of him, there is little to prevent him from achieving his goal. A little luck or some extra effort should be enough. Even if this way of thinking leads to verbal estimates that seem to violate normative rules of probability, it is based on causal considerations that make good pragmatic sense and are perhaps more adaptive, from both an individual and an evolutionary point of view, than the ability to calculate probabilities according to the 1/n rule.

Verbal expressions of probability appear to be of two kinds, directionally positive or negative, inviting framing effects. A medical treatment that offers "some possibility" of cure sounds more encouraging than a cure that is "quite uncertain," even if both phrases refer to the same level of probability (Teigen & Brun 1999). But such verbal frames may also facilitate normative probabilistic inferences. Conjunction errors, extensively discussed by S&W, apply chiefly to positive phrases, such as probabilities and possibilities. When one outcome is reported to be "somewhat probable" and the other has "a small probability," we found that their joint occurrence is given a modal rating of "somewhat probable," in line with the conjunction fallacy. But when outcomes are described negatively, in terms of uncertainties and doubts, it becomes easy to see, intuitively, that their conjunction must be *less* likely than the constituent events. Thus, the conjunction of a "quite uncertain" and a "somewhat uncertain" event was judged by most respondents to be "very uncertain" (Teigen & Brun 1999). In this case, intuitive thinking, triggered by a verbal frame, reduced the number of conjunction errors by 50%.

These observations in no way invalidate Stanovich & West's claim that people differ reliability in their analytic abilities and that high levels of analytic intelligence can play a major role in reducing the fundamental computational bias. At the same time, much theoretical and empirical work is needed to enable us to spell out the principles and abilities involved in successful intuitive thinking. Even when our analytic abilities fail, well-framed intuitions may come to our assistance in narrowing the normative/descriptive gap.

Beyond "pardonable errors by subjects and unpardonable ones by psychologists"

X. T. Wang

Psychology Department, University of South Dakota, Vermillion, SD 57069
xtwang@usd.edu

Abstract: Violations and biases relative to normative principles of rationality tend to occur when the structure of task environments is novel or the decision goals are in conflict. The two blades of bounded rationality, the structure of task environments and the computational capacities of the actor, can sharpen the conceptual distinctions between the sources of the normative and descriptive gap.

Rationality in the face of goal conflicts and compromise. Stanovich & West (S&W) have identified an under-examined and potentially very important topic: the implications of individual differences in reasoning for identifying the determinants of the observed gap between normatively defined rationality and human reasoning and decision making behaviors.

The four identified sources of the normative descriptive gap ostensibly fit, in a two by two manner, into the simple dichotomy of "pardonable errors by subjects and unpardonable ones by psychologists" (p. 340) that was criticized by Kahneman (1981). Beyond the simple dichotomy, S&W have mapped individual differences in SAT scores onto reasoning performance and two underlying reasoning systems. The first system appears to consist of a rich array of evolved, domain-specific mechanisms while the second one involves abstract, domain-general reasoning. It is likely that the second system is only a backup system for tasks with which the first system is not "programmed" to deal. As Simon (1956) suggests, "conflict of choice may often be equivalent to an absence of a choice mechanism in the given situation. And while it may be easy to create such situations in the laboratory, the absence of a mechanism to deal with them may simply reflect the fact that the organism seldom encounters equivalent situations in its natural environment" (p. 137). According to Simon, human rational behavior is shaped by a scissors whose two blades are "the structure of task environments and the computational capacities of the actor" (Simon 1990, p. 7).

Our recent studies have shown that violations of the descriptive invariance principle of utility theory appear and disappear as a function of the social structure of task environments. The occurrence of framing effects (i.e., the irrational reversal in risk preference due to descriptive frames of the same expected outcomes) entails two antecedent conditions where either (1) the task context is novel and lacks valid cues or (2) decision goals are in conflict (see the table below).

Table 1 (Wang). *Framing effects as a function of the structure of task environments, social cues, and decision goals (from Wang 1996; Wang et al., in press).*

Structure of Task Environments	Social Cues	Expected Survival Rate	Decision Goals	Choice	Framing Effect
Family (6 kin)	Kinship	1/3	Clear	Risk Seeking	No
Small Group (6 or 60 strangers)	Anonymity and Relatedness	1/3	Clear	Neutral or Slightly Risk Seeking	No
Mixed Group*	Kinship vs. Anonymity	1/3	in Conflict	Framing Dependent	Yes
Large Group (600, 6000, or 6 billion lives)	Anonymity and Evolutionary Novelty	1/3	Less Clear	Framing Dependent	Yes

*In the mixed group contexts, the endangered lives consist of either 1 kin and 5 strangers or 2 kin and 4 strangers.

In these experiments, subjects have to make a choice between a sure outcome saving one-third of the endangered lives and a gamble resulting in a one-third chance that everyone survives and a two-thirds chance that no one survives. As shown in the table, when a "kith-and-kin" rationality is activated, the majority of the subjects are risk-seeking under both framing conditions. However, irrational reversals in risk preference are found in the cases where either the structure of the task environments is evolutionarily novel (e.g., large anonymous groups) or decision goals are in conflict (e.g., mixed groups).

The two blades of bounded rationality, used as classifying devices, can sharpen the conceptual distinctions between the possible sources of the normative and descriptive gap. The first two identified sources can then be examined in terms of computational capacities while the last two may be better understood when the structure of task environments is taken into account.

Can a model of rationality ever achieve both normative adequacy and descriptive accuracy? S&W have depicted a picture of two camps fighting over the sovereignty of normative theories of rationality. Their intuitive theoretical proposal of two reasoning systems is unfortunately weakened by the loose Meliorist/Panglossian dichotomy which oversimplifies the diverse contributions made by each of the researchers classified in each camp. This dichotomy adds little to understanding the normative/descriptive gap but blurs the conceptual distinctions between theories of rationality.

For example, S&W cite Lola Lopes's argument twice to show that "a potent strategy for the Panglossian theorist to use against the advocate of Meliorism" is to argue that "the gap between the descriptive and normative occurs because psychologists are applying the wrong normative model to the situation" (sect. 4, para. 1). However, in fact, Lopes (e.g., 1984; 1987; 1990) was among the first to provide solid empirical evidence that individuals differ in the relative emphasis they put on the security level versus the potential of risky alternatives and that such differences affect their choices. Lopes's security-potential and aspiration level (SP/A) theory provides a psychological mechanism for the effects not only of situational variables but also of individual dispositions on choice behavior.

While the theory strives for descriptive accuracy, it is also formulated using normatively adequate rank-dependent utility functions. Rank-dependent utility models reflect an attempt to achieve the nonlinearity in decision weights necessary to account for people's deviations from expected utility theory and, at the same time, eliminate theoretical violation of stochastic dominance. The axiom of stochastic dominance serves as a cornerstone of normative models of rationality and holds empirically as well or is violated only in very special cases (see also Luce 1991; 1992; Quiggin 1982; Weber 1994; Yaari 1987).

Rational principles and axioms have been used as benchmarks in the descriptive, prescriptive, and normative models for gauging reasoning and decision making behaviors. Without these common benchmarks, the notions of "deviation," "violation," "error," "illusion," "fallacy," and "bias" would become theoretically meaningless. For a similar token, Gigerenzer and Hoffrage's (1995) work on how to improve Bayesian reasoning can be seen as a further validation of Bayesian inferences. As a rational principle, Bayes's theorem will hold descriptively as long as its required information is accessible and provided in an intuitive format (e.g., natural frequency). Reasoning and decision making models should strive to achieve both descriptive accuracy and socially and ecologically defined normality by taking into account the constraints of bounded rationality.

Implicit learning of (boundedly) rational behaviour

Daniel John Zizzo

Brasenose College, Oxford University, Oxford OX1 4AJ, United Kingdom
daniel.zizzo@hertford.ox.ac.uk

Abstract: Stanovich & West's target article undervalues the power of implicit learning (particularly reinforcement learning). Implicit learning may allow the learning of more rational responses – and sometimes even generalisation of knowledge – in contexts where explicit, abstract knowledge proves only of limited value, such as for economic decision-making. Four other comments are made.

Stanovich & West (S&W) have produced a stimulating target article with many valuable points, from their review of the importance of systematic errors to that of the correlation between intelligence measures and experimental performance.

S&W propose a dual process theory of human reasoning. System 1 (S1) reasoning is associative, implicit, highly contextualized, and characterised by a fundamental computational bias. System 2 (S2) reasoning is rule-based, explicit, abstracting, and entailing a utility optimising choice.

In Table 3, S&W recognise that S1 may not be simply a genetic universal; it may also be acquired by "exposure and personal experience." Nevertheless, in the main text S&W focus only on genetic sources, producing ubiquitous computational biases. To the extent that such biases are *not* universal, this is attributed to individual differences in S2 cognitive abilities. S&W believe that humans can become more rational (and should do so in today's "competitive capitalist economy"), and that the way to attain this is to become more capable of abstract S2 reasoning.

S&W's S1 corresponds to the implicit learning route discussed by Shanks and St. John (1994). But S&W make no reference to anything like implicit learning in the text (unlike the table). If they did, they might be forced to recognise that there can be individual differences in knowledge and heuristics used, due not to variance in S2 abilities, but simply to differences in the past learning histories of the subjects. This may be sufficient, for example, for subjects to learn to commit or avoid the base rate error (Goodie & Fantino 1996). Subjects in economic experiments with monetary incentives may sometimes acquire a knowledge of general strategies and rules if given purely behavioral reinforcement (Stahl 1996; Warnick & Slonim 1999). Neural network models show that the reinforcement learner may generalise a capability to behave in a reasonably (i.e., boundedly) rational way in a variety of cases. Outside traditional cases (e.g., language learning), this may include the learning of (boundedly) rational behaviour in interpersonal (e.g., game theoretical) economic problems. In Zizzo (1999), neural networks "learn" by repeated exposure to examples – arguably just like consumers, workers, firms, and other agents. The predictions of the model of learning-to-be-rational are shown to fit experimental data. S1 may be "highly contextualised," but generalisation *is* possible.

Does this matter? It does, because otherwise we are left with the only alternative option of teaching logic on a mass scale. Between a generalised application of behavioural reinforcement and the option of having logic classes for everyone, the truth for any cognitive engineer should probably lie in the middle. Teaching logic may be relatively ineffective, as in the case of the conjunction fallacy (Stolarz-Fantino et al. 1996). Also, there are arenas where behaviour is largely guided by implicit knowledge: *economic decision-making is one of them.* Very few economists would think that agents are explicitly optimising their utility function subject to their budget constraint when buying baked beans at the supermarket: even if they wanted to do so, they would be unable to, because it would require a knowledge of calculus. Most economic decisions are more complex than buying baked beans: to make the applicability of economic rationality depend on an explicit and conscious rational choice is equivalent to making rational choice inapplicable except perhaps to a tiny minority of mathematically inclined economists. Rather, rationality in economic decision-making is linked to implicit knowledge. Experimental evidence shows that reinforcement learning plays an important role in learning (or failing to learn) more rational responses (e.g., Roth & Erev 1995), together possibly with initial predispositions (Camerer & Ho 1999). Moreover, as very few agents (mostly economists) have explicit knowledge of the optimal choice in many cases, this is a field where, if we were to believe S&W's argument about the universality of the fundamental S1 bias, we should not observe systematic differences in degrees of rationality across subjects: however, such differences clearly exist (e.g., Stahl 1998).

Other points:

1. *Reinforcement learning may be maladaptive.* In section 6.2, S&W correctly suggest that agents may not have an evolutionarily optimal utility function, and that this may be relevant for S2 cognition. Nevertheless, the same arguments apply to S1. There is no reason to believe that implicit learning is not flexible enough to produce outcomes that are potentially maladaptive genetically. Joe is positively reinforced by having sex with condoms and feels "safe," therefore he keeps using condoms to have sex.

2. *The conjunction fallacy may not be purely linguistic:* Fantino and Savastano (1996) replicated the conjunction fallacy using a purely behavioural experimental paradigm, although they found a tendency to add single event probabilities rather than to average them out as we did (Zizzo et al. 2000). In a recent experiment I used a similar purely behavioural setting to test the effect of different reinforcement learning histories on the conjunction fallacy and the use of heuristics.

3. *Rational agents must often take random performance errors into account.* Such errors are *not* irrelevant for economic rational decision-making in many settings, such as market entry deterrence (Kreps & Wilson 1982) and stock markets (DeLong et al. 1990). Selten (1975) and recently McKelvey and Palfrey (1995; 1998) have introduced new equilibrium concepts to analyse the effect of random trembles (i.e., perturbations induced by mistakes) on game equilibria.

4. *Rational agents may be rational fools* (Sen 1977): In social dilemmas (where it is individually optimal to defect but socially optimal to contribute to the public good), "rational fools" knowing the "rational" solution end up worse off – socially *and* individually – than dumb cooperators playing with one other (for finitely repeated Prisoner's Dilemmas, see Neyman 1985). Economists, who receive explicit training in game theory, typically end up worse off than noneconomists (Frank et al. 1993). The implication for social policy is that there may be vital contexts where what is normatively rational should not be made a matter of sheer logical beauty: "interactive intelligence" deserves its due.

Authors' Response

Advancing the rationality debate

Keith E. Stanovich[a] and Richard F. West[b]
[a]*Department of Human Development and Applied Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1V6*
[b]*School of Psychology, James Madison University, Harrisonburg, VA 22807*
kstanovich@oise.utoronto.ca
westrf@jmu.edu falcon.jmu.edu//~westrf

Abstract: In this response, we clarify several misunderstandings of the understanding/acceptance principle and defend our specific operationalization of that principle. We reiterate the importance of addressing the problem of rational task construal and we elaborate the notion of computational limitations contained in our target article. Our concept of thinking dispositions as variable intentional-level styles of epistemic and behavioral regulation is explained, as is its relation to the rationality debate. Many of the suggestions of the commentators for elaborating two-process models are easily integrated into our generic dual-process account. We further explicate how we view the relation between System 1 and System 2 and evolutionary and normative rationality. We clarify our attempt to fuse the contributions of the cognitive ecologists with the insights of the original heuristics and biases researchers.

Our target article had two major themes, and commentators tended to focus on one or the other. The first of our themes was to demonstrate how patterns of individual differences can be used to make inferences about the reasons that normative models and descriptive models may not coincide. Secondly, we sketched a generic two-process model and related it to the concepts of normative and evolutionary rationality in order to reconcile the findings and theories of the cognitive ecologists with those in the original heuristics and biases literature. We turn first to the critiques of these two major themes and then deal with clusters of other issues.

R1. Individual differences and the normative/descriptive gap

As we fully expected, there was considerable dispute about using individual differences as a clue to the validity of the wrong norm application or alternative construal explanations for normative/descriptive gaps. We discuss these more controversial applications first and then elaborate on some criticisms of our discussion of computational limitations and performance errors.

R1.1. Normative applications versus normative models.

Several commentators (e.g., **Ayton, Baron, Frisch, Goodie & Williams, Hardman, Krueger, Schneider**) seem to have understood us as claiming that patterns of covariance in performance can help to justify normative models *themselves,* when in fact our argument amounted to the weaker claim that patterns of individual differences can help to determine whether it is appropriate to *apply* a particular normative model to a specific problem. We had tried to clarify this difference during the review process for the article, but, given the number of commentators who misinterpreted us, we were obviously unsuccessful. We thus agree with commentators such as Baron, Frisch, and Krueger that the data we present do not bear on the validity of normative models themselves. Nevertheless, there remain the many disputes about which normative model should be applied to a particular experimental situation. In the same essay quoted in section 5.1 of the target article, Tversky (1975) emphasized that this is often the case:

> Furthermore, the results raise serious questions regarding the interpretation of consequences or outcomes, and highlight some of the difficulties involved in the normative *application* of utility theory. . . . Proponents and opponents of utility theory argue about the validity of certain axioms (e.g., substitutability, independence) where in fact the *applicability* of the axioms and not their adequacy is at stake (pp. 163, 172, emphasis added).

That empirical data may well be relevant to *these* disputes – highlighted in the Tversky quote – is the heart of our argument.

Again, to clarify a point we insufficiently stressed, we are not attempting to explicate the normative standards themselves with our methods. Instead we are attempting to add an empirical leavening to disputes about appropriate normative applications 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). For example, the issue is not whether or not there is something wrong with Bayes' rule as a normative model. The issue is whether the rule is appropriate to apply to a particular thought-problem presented to a subject. Consider the notorious AIDS and cabs problems discussed in the target article. A host of philosophers and others have questioned the application of Bayes' theorem to these problems. They argue that because the base rate lacks reference class specificity its use is negated in these problems. The critics are not arguing against Bayes' rule *in situations where it applies.* Instead, they are arguing that it is inappropriately applied in these cases.

The controversies about the results in the heuristics and biases literature more often concern the application of normative models rather than the validity of certain normative models in the abstract. As explained in section 5 of the target article, alternative task construal is a type of incorrect norm application that results not from the experimenter invoking the wrong model for the problem as *set* (as in the false consensus effect), but of the experimenter invoking the wrong normative model because the subject has interpreted the problem differently and is answering a functionally different question.

Although we agree that the data in the target article and in Stanovich (1999) do not help to discover or validate norms themselves, they do at least provide weak diagnostic information about whether we have applied the right normative model to a particular situation – and this problem has been at the heart of the debates in the heuristics and biases literature. Furthermore, as argued in section R1.3, it is a necessary task to undertake – if we do not, we risk slipping into the solipsistic relativism that **Manktelow** refers to at the end of his commentary.

Several commentators, most notably **Oaksford & Sellen,** interpreted our methods exactly as we intended. These commentators describe an interesting study demonstrating that high schizotypy was associated with *more* logical performance on a conditional inference task. They use this finding to infer, as we would, that logic is not the appropriate normative model of conditional inference for the particular task that they used.

R1.2. The understanding/acceptance assumption.

Although many commentators either explicitly or implicitly accepted our use of the understanding/acceptance principle (e.g., **Kahneman, Klaczynski, Kühberger, McCain, Oaksford & Sellen, Okasha, Over & Evans**), others (e.g., **Ayton, Hardman, Schneider**) question the usefulness of the understanding/acceptance assumption as a tool for adjudicating different explanations for the normative/descriptive gap. Schneider argues that the understanding/acceptance assumption should be viewed as a necessary but not sufficient criterion. We are in sympathy with this view because we have always stressed that it is one of many (probably weak) empirical indicators that might be of use in resolving these vexing disputes about human rationality.

We are in less sympathy with **Schneider**'s critique of Larrick et al's. (1993) articulation of the understanding/acceptance assumption because Larrick et al.'s view allegedly has the implication that all intelligent people will be in agreement on the best course of action. Here, Schneider, like **Hardman,** appears to apply the assumption in an overly deterministic manner – as in the latter's comment that some people do not necessarily accept a decision-theoretic norm that they understand. But, this is a case of setting up a strawman so that it can be easily knocked down. There are *no* such deterministic principles in all of psychology. The assumption, as we and others (e.g., Larrick et al. 1993) have applied it, is – like all others in psychology – a probabilistic inference: that the correct normative application is *more likely* to be invoked by more intelligent people. Schneider characterizes the idea "that intelligent people will be in agreement on the best course of action" as "oversimplified," but the notion is not at all oversimplified when interpreted as a standard probabilistic contingency rather than as an unrealistically deterministic rule. In fact, in a study of informal reasoning (Stanovich & West 1997) we found that untrained students of high (but not extraordinarily high) cognitive ability tended to agree more with a group of experts (philosophy professors) who evaluated the same informal arguments than they agreed with their student peers of lower cognitive ability.

Hardman is right that the understanding/acceptance principle is most discernible in the dialogue character Dr. S. (Savage) in the Slovic and Tversky (1974) paper. This is why, after citing Larrick et al.'s (1993) transparent application of the principle ("intelligent people should be more likely to use the most effective reasoning strategies," p. 333), we were careful to point out that "Slovic and Tversky

(1974) made essentially this argument years ago, although it was couched in very different terms in their paper." Regardless of whether or not Slovic and Tversky themselves endorsed the principle, the understanding/acceptance assumption arises not uncommonly in the debates about rationality in the heuristics and biases literature. As we were at pains to show in section 4.3, the principle has been tacitly invoked by all sides in the rationality debate, and it has been used in various arguments by investigators with perspectives as diverse as Cohen (1982), Funder (1987), Larrick et al. (1993), Lopes and Oden (1991), and Wetherick (1971).

However, both **Ayton** and **Hardman** make the wrong inference from the fact that in the Slovic and Tversky (1974) study an application of the understanding/acceptance assumption does not resolve the debate in favor of the appropriateness of Savage's independence axiom for the Allais problem. As we ourselves pointed out in the target article in section 4.2, when presented with arguments to explicate both the Allais (1953) and Savage (1954) positions, subjects found the Allais argument against independence at least as compelling and did not tend to change their behavior in the normative direction. Here we have a link with the issue just discussed in section R1.1. Ayton and Hardman both seem to be invoking the understanding/acceptance principle to adjudicate a normative rule *itself* (in this case, the independence axiom), and, because they view the rule as uncontroversial, the failure of the understanding/acceptance principle to converge upon it is seen as a defect in the principle. But as discussed by Schick (1987), Broome (1990), and Tversky (1975), the arguments presented to the subjects in the Slovic and Tversky (1974) paper are not about the axiom per se – they are about the *appropriateness* of applying it in this particular situation. Tversky (1975) himself was clear that the Allais paradox concerned the issue of norm application and not the independence axiom itself: "the key issue [in the Allais paradox] is not the normative adequacy of the independence principle, but rather the legitimacy of various interpretations of the outcomes, and this issue lies outside the scope of utility theory" (p. 170).

What is at issue is two different task construals – one which codes regret into options and one which does not. Indeed, the argument attributed to Allais in the Slovic and Tversky (1974) paper was couched largely in terms of regret (correctly, according to Lopes, 1988, who argued that "Allais offered his problems to illustrate the operation of psychological mechanisms that are disallowed by both classical and modern expected utility theory," p. 405). However, as Schick (1987) and Broome (1990) have pointed out, there is no violation of the independence axiom in the traditional Allais choice as long as we allow regret to be coded into outcomes (see Tversky 1975). Coding of regret into the outcomes is a much more controversial proposition than the independence axiom itself. Thus, the failure of the understanding/acceptance principle to adjudicate this particular dispute does not in any way undermine the principle – it is in fact consistent with the vexatious status of this problem in the decision theory literature.

The other important lesson taught by this discussion of the Allais paradox is that the issue of rational task construal is not going to disappear. We would issue this caution to those investigators who (perhaps because of the confusion discussed in sect. R1.1) would prematurely jettison the understanding/acceptance assumption as a crude method for adjudicating disputes about the reasons for normative/descriptive gaps. The problem of determining rational task construal will continue to loom large in psychological studies of judgment and decision making and, in the absence of formal arguments, we will need to make use of any relevant information available. The alternative, as we argued in section 5.1., is a relativism that the field has not fully thought through.

R1.3. Relativism and rational task construal. Most people share two intuitions that are in sharp conflict. Most of us are not complete relativists – we feel that there are at least *some* principles of cognitive evaluation. But there is another intuition that is completely at odds with this one. It is the widespread view (see Henle 1962; 1978) that all subject interpretations of a problem – even the most bizarre – must be honored and treated equally. The problem is that the latter view is in fact sharply at odds with the notion that there are *any* rational standards. The difficulty is that if we accept the Panglossian assumption that there is no construal so implausible that it should not be considered rational, then even an inconsistency as extreme as intransitivity can be neutered by concocting a construal of the problem that removes the intransitivity. Following the example discussed by Broome (1990) that was mentioned in the target article, pretend that Bill prefers, successively, object A to object B, B to C, and then C to A – apparent intransitivity. We might posit that, to Bill, the third choice involved not "A" but instead something that he puts a different value on: "object A offered immediately after a choice involving two things that are not-A." And to him, this bizarre entity – "object A offered immediately after a choice involving two things that are not-A" – is not valued the same as "object A offered immediately after a choice involving only one thing that is not-A." Thus, an "object A offered immediately after a choice involving two things that are not-A" might as well be designated D – and there is no inconsistency at all in preferring A to B, B to C, and then C to D. Bill is now no longer intransitive.

However, despite his perfect transitivity, despite his perfectly rational competence, Bill is still a money pump – and it is very doubtful that his preferences are serving his goals (Baron 1993; 1994). Nevertheless, Panglossians still endow Bill with perfect rational competence, with perfect transitivity. But this perfect rational competence that Bill enjoys is utterly uninteresting (and unhelpful to him in the real world in this case – because he is a money pump). The only interesting thing here is why Bill construes the situation in this bizarre way. The several commentators who emphasized the consequences of actions in the real world (e.g., **Frisch, Funder, Hardman, Schneider**) would presumably be concerned about this money-pump implication.

Thus, the Meliorist might have begun this inquiry with the question: "Why isn't Bill transitive?" The Panglossian comes along and says, "But you're wrong – he is! Here is a task construal that makes his choice rational." We can oblige the Panglossian in this relativist way, but then the interesting question (and it has become no less interesting) is why Bill construes the situation in this weird way. The Panglossian is happy because the question has been transferred from the competence side of the ledger (which gets the answer the Panglossian wants – that Bill is rational) to the performance side. But if the Panglossians get their way, *all* of the interesting questions about performance variability will end up on this side.

Adler, in his commentary, and Margolis (1987) make just this argument – that alternative task construals can be used to insulate reasoning competence from charges of irrationality, but that this ploy simply transfers the irrationality to the stage of problem representation. Irrationality has not been abolished, it simply has been transferred to a different cognitive operation. Margolis (1987) argues that because for many tasks in the heuristics and biases literature "the alternative interpretation necessary to make the usual response logical is either not there at all, or there only under some interpretation so tendentious as to transfer the puzzle about the logic of the response undiminished to an earlier stage" (p. 20). Likewise, Ader, in his commentary, points out that the contextualized interpretations may reflect a defective grasp of important logical terms.

The problem for the Panglossian is that they have only one degree of freedom – they must posit a construal that makes choices rational. This of course creates problems when the normative/descriptive gap is quite large. If the competence model is fixed at perfect rationality, the only way to explain an extremely large normative/descriptive gap is by an equally large deviation from the normal construal of the problem. Bar-Hillel (1991) has argued that

> many writers have attempted to defend seemingly erroneous responses by offering interpretations of subjects' reasoning that rationalizes their responses. Sometimes, however, this charitable approach has been misguided, either because the subjects are quick to acknowledge their error themselves once it is pointed out to them, or because the interpretation required to justify the response is even more embarrassing than the error it seeks to excuse (p. 413).

The Meliorists are not forced into positing task construals "even more embarrassing than the error they seek to excuse" because they have another mechanism for explaining such gaps – the subject's reasoning processes can be viewed as deviating from principles of rational thought. By positing some deviations from rational principles, the Meliorist is not trapped into offering bizarre task construals in order to account for the discrepant behavior. Using the understanding/acceptance principle as a clue to infer whether some task construals are more rational than others is a way of avoiding what **Manktelow** refers to as the trapdoor of relativism.

The importance of avoiding this trapdoor can perhaps be better appreciated by considering an analogous trapdoor in the moral domain. Note that one could assure perfection in the moral domain by the expedient of honoring all judgments of moral relevance. Imagine that people were allowed to say "no, that situation is not in the moral domain" when we attempted to evaluate their moral judgments and behavior. If people were allowed to exclude from the moral domain whatever situations they wanted, and then proceeded to exclude all contentious ones, we could never judge them as immoral. But surely something has gone wrong here. Recognizing the situation as having moral dimensions is itself a large part of morality – just as the construal of a problem is a large part of rationality. Just as we would not want to set aside all moral evaluation because we too permissively accepted all judgments of moral relevance, we would not want to set aside all cognitive evaluation because we too permissively accepted all task construals as rational.

R1.4. Cognitive ability and the SAT. Some commentators (e.g., **Goodie & Williams, Jou, Manktelow, Sternberg**) claim to see some circularity in the correlations between cognitive ability measures such as the SAT and certain tasks from the heuristics and biases literature. Others see the correlations as "not surprising" (**Hunt, Jou**). But it is fundamentally mistaken to see these correlations merely as instances of "one reasoning task correlating with another." As we argued in the target article – and as is clear to anyone who immerses themselves in the controversies surrounding the heuristics and biases research program – scoring a vocabulary item on a test from the educational psychology literature and scoring a probabilistic reasoning response are not the same. The normative appropriateness of responses on tasks from the latter domain has been the subject of extremely contentious conceptual dispute in a way that the former responses have not.

The heuristics and biases literature 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 tasks we examined were vastly more varied than this. The choice between a vivid case and a statistical fact on a Nisbett-type problem is nothing like an item on the SAT; neither is the calibration curve indicating an overconfidence effect in a knowledge calibration experiment; neither is the combining of a diagnostic indicator and base-rate information; neither is the assessment of the tendency to honor sunk costs; and so on.

This point is clearly recognized by **Oaksford & Sellen** who reinforce it by noting that

> moreover, 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.

Thus, it is not the case that the judgment and decision making tasks in the heuristics and biases literature are "just another way of measuring intelligence" as some commentators (e.g., **Jou**) seem to imply. Neither an analysis of the tasks themselves, nor a consideration of the full range of our results (that is, beyond those displayed in Table 1; see Stanovich 1999) support such a view. Just because two tasks fall under a generic category such as "reasoning" – the type of category used in survey textbooks – does not make them measures of the "same thing."

What some commentators seem to have missed is that – as demonstrated by **Oaksford & Sellen** and in our analysis of the false consensus effect, overconfidence effect, and noncausal base rates – the correlations *are not always positive* (several commentators *did* note this important fact, see **Goodie & Williams, Krueger, Manktelow**). As we argued in section 4.4, the very obviousness of positive manifold is what makes violations of it interesting. Even the critical commentators agree that hundreds of studies in the psychometric tradition make positive manifold an obvious default assumption – one, we argued, that can be used as a marker for the appropriate model to apply and for the appropriate task construal.

In short, 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. In fact, the magnitude of the normative/descriptive gap displays all manner of correlations, from negative through zero to positive, with cognitive ability. We posited that the reason the magnitude of this gap has drastically different correlations with the ability to reason on other tasks is that the *cause* of the gap is different across different tasks in the heuristics and biases literature.

Some commentators raised issues related to our use of cognitive ability tests like the SAT – issues such as the implications of group factors of intelligence and the tight relationship between general intelligence and working memory that we assumed (**Epstein; Klaczynski; Sternberg**). On the latter, we note that **Bucciarelli** agrees with our emphasis on the importance of working memory as a key indicator of computational capacity that relates to performance on tasks such as syllogistic reasoning (see also, Bara et al. 1995; Barrouillet & Lecas 1999). Furthermore, we would draw attention to a recent paper by Engle et al. (1999) where, based on extensive structural equation modeling of memory tasks, a strong linkage between fluid intelligence and working memory was found – consistent with the literature we relied on in the target article (Kyllonen 1996; Kyllonen & Christal 1990). Additionally, Engle et al. (1999) found a substantial linkage between working memory and *both* Math SAT and Verbal SAT scores. Finally, we note that the executive, inhibitory, and planning functions of working memory that relate most closely to fluid intelligence (Duncan et al. 1996; Engle et al. 1999) are System 2 processes – not the more System 1-like structures such as the articulatory loop and visual store as suggested by **MacDonald & Geary.**

Regarding SAT and education (**Schneider**), see Table 2.3 of Stanovich (1999) for evidence that the correlation between performance on heuristics and biases tasks and training in mathematics and statistics was negligible and much lower than the correlations with cognitive ability. The latter correlation is not simply the result of differential educational experience.

On the issue of group factors, in Note 4 we drew attention to the fact that, with our practice of focusing on the SAT total score, we did not wish to imply the denial of the existence of second-order factors in a hierarchical model of intelligence – which we view as an established fact (Carroll 1993). Nevertheless, whether a theorist emphasizes stratum III of Carroll's (1939) model (g) or stratum II (group factors), all are factoring the same positive manifold – and *violations* of positive manifold are equally surprising from whatever factor analytic perspective is driving the analysis.

Two further empirical findings are relevant. First, separating out the verbal and mathematical scores on the SAT reveals virtually identical trends in our studies as does analyzing only the total score. Secondly, as noted in section 3, in virtually all of our studies, we employed converging measures of cognitive ability – usually an additional test of crystallized intelligence such as a vocabulary measure and a measure of fluid intelligence (usually the Raven matrices). The correlations obtained with these measures converge with those obtained with the SAT. The use of these alternative measures also relates to the issue of circularity raised in the **Goodie & Williams** and **Sternberg** commentaries that we mentioned above: There is nothing circular about using a checklist measure of vocabulary to predict probabilistic reasoning in the Linda problem.

Finally, as we noted, there are good reasons to view such cognitive ability tests as crude indicators of the overall level of current computational efficiency (despite the fact that the latter is undoubtedly the result of many individual subprocesses, see Hunt 1999). First, as mentioned above, there is the strong link with working memory (Engle et al. 1999; Kyllonen 1996) – the quintessential indicator of computational capacity in cognitive science. Secondly, there is the link with the crucial planning, attentional, and executive functions of the frontal lobes (Duncan et al. 1995; 1996; Engle et al. 1999; Hunt 1999; Pennington & Ozonoff 1996). Thirdly, there is the work on the links with neurophysiological and information processing indicators of efficient cognitive computation that we cited in the target article.

R1.5. Alternative interpretations of computational limitations. The commentaries by **Funder** and **Jou** prompted some further thoughts on the tendency in cognitive science to excuse performance suboptimalities when computational limitations have been demonstrated (additionally, see Ch. 8 of Stanovich [1999] where computational limitations are treated in a more nuanced way than is apparent in the target article). First, we agree completely with **Funder** and **Krueger** that the correlations in Table 1 are no doubt attenuated due to the limited number of trials in most of the heuristics and biases tasks. The disattenuated correlations are obviously higher, and we await psychometrically more rigorous work than ours for more accurate measures of the true magnitude.

Funder is right that we do not completely share his pessimistic conclusion regarding Meliorism – but this is mainly for reasons not stated in the target article. There, we adopted the most conservative position (relative to Meliorism) regarding the interpretation of computational limitations. A high disattenuated correlation between cognitive ability and a rational judgment task is only discouraging for Meliorism if a computational limitation is viewed as a fixed absolute. However, in the human case, this is too strong an assumption – one that is overly stacked against the Meliorist view. Computational limitations in humans are not fixed in the way they are in a computer (and even there complications arise in the case of neural networks undergoing learning). The computational limits of humans are not as precisely quantifiable as those of nonhuman machines, and the consensus view of developmental psychologists is that they should not be conceived as absolute limitations (Ceci 1996; Neisser et al. 1996; Perkins & Grotzer 1997). A host of experiential variables have been shown to affect the cognitive ability of humans, particularly in early developmental periods but not limited to these periods (Ceci & Williams 1997; Cunningham & Stanovich 1997; Keating & Hertzman 1999; Morrison 1997; Myerson et al. 1998; Nickerson 1988; Stanovich 1993).

Also, in section 7 of the target article, we alluded to work showing that thinking dispositions that are related to epistemic regulation can also predict performance on a variety of rational thinking tasks, and to some extent they can predict performance independently of cognitive ability. In section R1.7 we will discuss this finding – relevant to the comments of several commentators – in more detail because it was given short shrift in the target article. If, as some theorists (e.g., Baron 1985a) argue, thinking dispositions are more malleable than cognitive ability, this again provides a lever for attempts to make thinking more normatively rational. Additionally, if we accept views in which normative rationality and morality largely coincide (Gauthier 1986), then it is clear that many cultural institutions (parents, schools, churches, etc.) have as a practical goal the teaching of rationality. These efforts make sense only if cognitive limitations are not fixed and if we view normative competence in the reasoning domain as at least in part a cultural product.

More important, in the target article, the dominant interpretation of cognitive ability measures that we utilized was to view them as omnibus indexes of the efficiency of computational components such as working memory, information retrieval speed, and so on – that is, as proxies for overall cognitive power. However, an alternative interpretation might highlight instead the metacognitive functions of intelligence (components having to do with flexible strategy use, see Sternberg 1985) rather than the elementary information processing components. This alternative conceptualization views cognitive ability measures not as indicators of basic computational power but as indicators of the tendency to respond to problems with appropriate strategies. These alternative notions of cognitive ability have waxed and waned throughout a century of work on human intelligence (Sternberg 1990; Sternberg & Detterman 1986).

Under the first interpretation it is assumed that positive correlations result because normative responses are computationally more complex, and only those people with the requisite computational power are able to compute them. Alternatively, under the second interpretation of cognitive ability, the normative strategy might *not* be more computationally complex. It might simply be more efficient and more readily recognized as such by more metacognitively sensitive individuals. This second interpretation was largely ignored in the target article because it makes use of metacognitive notions that blur the line between the intentional and algorithmic levels of analysis – a distinction that proved useful in framing the results of many of the experiments.

Finally, the cognitive ecologists stress a different type of computational limitation that highlights another way of characterizing overall cognitive ability. From within this framework, computational limitations result from problems being presented to the brain in formats that are incongruent with the representational format used by the relevant cognitive module (Brase et al. 1998; Cosmides & Tooby 1996; Gigerenzer & Hoffrage 1995). Cognitive ability might therefore be characterized as the number of different problem representations with which the individual can cope. The Meliorist – free of the nativist assumptions of many cognitive ecologists and placing more emphasis on the cultural malleability of problem representations – would be more likely to see variation in this quantity. Thus, under this representational view, correlations between cognitive ability and the normative strategies could be interpreted as indicating that individuals who are higher in cognitive ability have more alternative problem representations and thus would be more likely to have available to them the representation that is appropriate for the way in which the experimenter has framed the problem.

In short, moving outside of the technical area of computational complexity theory (see Oaksford & Chater 1992; 1993; 1995) into the looser vocabulary of human abilities research, it might be possible to distinguish three different types of computational limitations: Type A: limitations in the efficiency of basic information processing components; Type B: limitations in metacognitive control (in the flexible deployment of strategies to solve a problem); and Type C: limitations in the types of representations that the cognitive apparatus can deal with and the types of procedures it can implement.

R1.6. Alternative construal as a computational escape hatch. We are very sympathetic to **Ball & Quayle**'s development of the idea of computational escape hatches – an idea we discussed in the context of alternative task construals in Note 7 and in Stanovich (1999). There, we suggested that the notions of alternative task construal and computational limitations as explanations of the normative/descriptive gap might not be as separable as we implied. Our idea was that sometimes an alternative construal might be hiding an inability to compute the normative model. **Adler** suggests something similar when he argues that subjects sometimes make task interpretations consistent with conversational principles because they do not have on-line grasp of important distinctions.

Alternative construals as computational escape hatches could come about in one of two ways – either with or without metacognitive awareness (note that **Ball & Quayle** emphasize the former). In the first, the subject is aware of alternative task interpretations and chooses the one with the lowest computational demands. In the second, the computational escape hatch is used automatically and without awareness of the alternative interpretations (in short, in the manner of a System 1 module), perhaps because alternative representations on which to map the problem do not exist (i.e., there is a Type-C limitation). If the former – if subjects are actually choosing different task construals based in part on computational considerations – then such a mechanism may allow for the influence of thinking dispositions on performance. For example, in arguing that it is unlikely that a descriptive model will exactly mirror a normative one, Shafir (1993) argues that "to suppose that a single theory can serve as both a normative guide and a descriptive model of decision making requires that the cognitive processes that guide behaviour conform with our desire to reach rational decisions" (pp. 260–61). "Our desire to reach rational decisions" – clearly captured in the need for cognition, need for closure, and other dispositional constructs studied by ourselves and other investigators (see commentaries by **Epstein, Klaczynski,** and **Newstead;** and also Sternberg 1997c) – may vary between individuals and may be why one individual chooses a normatively rational but computationally costly task construal when an alternative is available.

R1.7. Thinking dispositions and cognitive styles. Such an emphasis on thinking dispositions and cognitive styles is urged by several commentators (e.g., **Epstein, Friedrich, Girotto, Klaczynski, Kühberger, Newstead, Oaksford & Sellen**) and was given rather short shrift in section 7 of the target article. In fact, in several papers (Sá et al. 1999; Stanovich 1999; Stanovich & West 1997; 1998c) we have examined a variety of these dimensions including some of the thinking styles alluded to by the commentators (e.g., reflectivity/impulsivity, need for cognition, vigilance, categorical thinking, flexible thinking, counterfactual thinking, and actively open-minded thinking). We completely agree with **Kühberger** that being higher in cognitive capacity is not equivalent to investing more cognitive capacity, which is why we have examined the need-for-cognition variable in recent studies (Stanovich & West 1999) and why we cited Smith and Levin's (1996) examination of this variable when discussing the understanding/acceptance assumption. The very point made by Kühberger is what has motivated our search for capacity-independent variance in thinking dispositions.

We have proposed that thinking dispositions should be distinguished from cognitive capacities because the two constructs are at different levels of analysis in cognitive theory and do separate explanatory work. Recall from section 3 of the target article that each level of analysis in cognitive theory frames a somewhat different issue. At the algorithmic level the key issue is one of computational efficiency, whereas issues of rationality arise at the intentional level. Using this taxonomy, we have proposed that omnibus measures of cognitive ability such as intelligence tests are best understood as indexing individual differences in the efficiency of processing at the algorithmic level. In contrast, thinking dispositions, as traditionally studied in psychology (e.g., Cacioppo et al. 1996; Kardash & Scholes 1996; Klaczynski et al. 1997; Kruglanski & Webster 1996; Sá et al. 1999; Schommer 1990; 1993; 1994; Stanovich & West 1997; Sternberg 1997c) 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. For example, in his model of mind as a control system, Sloman (1993) views desires as control states that can either produce behavior directly or through a complex control hierarchy by changing intermediate desire-states. He views dispositions (high-level attitudes, ideals, and personality traits) as long-term desire states that "work through a control hierarchy, for instance, by changing other desire-like states rather than triggering behaviour" (p. 85). It is through such a notion of superordinate and subordinate goal states that we might begin to get a grip on the vexing issue raised by **Manktelow** – that of assessing the normative status of the weighting given to short-term and long-term goals (Ainslie 1992; Baron 1993; 1998; Bratman et al. 1991; Haslam & Baron 1994; Nathanson 1994; Parfit 1984).

Thus, thinking dispositions are reflective of intentional-level psychological structure. It has been the goal of our research program to determine whether such features of intentional-level psychology can serve as explanatory mechanisms in accounts of discrepancies between normative and descriptive models of behavior (Stanovich 1999). If thinking dispositions correlate with individual differences in the normative/descriptive gap, then this will be prima facie evidence that the gap is caused by real differences in intentional-level psychology. However, any such association might well arise because the variation in thinking dispositions is co-extensive with differences in computational capacity. Thus, we thought it important to examine whether intentional-level cognitive dispositions can explain unique variance – variance independent of cognitive capacity. This has been one of the major analytic tests that we have used when examining individual differences across a variety of rational thinking tasks in the heuristics and biases literature.

We have consistently found (see Stanovich & West 1997; 1998c; Sá et al. 1999) that, *even after controlling for cognitive ability,* individual differences in performance on a variety of reasoning and decision making tasks can be predicted by measures of several of the thinking dispositions mentioned above. These findings converge with those of other investigators (and several commentators such as **Epstein, Klaczynski,** and **Newstead**) in supporting a conceptualization of human cognition that emphasizes the potential separability of cognitive capacities and thinking styles/dispositions as predictors of reasoning skill (e.g., Baron 1985a; Ennis 1987; Kardash & Scholes 1996; Klaczynski et al. 1997; Norris 1992; Schommer 1990; Sternberg 1997c).

Finding variance in intentional-level functioning that is not explained by computational capacity is an issue of considerable interest in philosophy and cognitive science because, as discussed in Stanovich (1999), there are three powerful traditions in philosophy that argue against the possibility of actual (as opposed to apparent) variation in the optimality of intentional-level psychologies. Arguments from charity (Dennett 1987; Quine 1960; Stein 1996; Stich 1990), from reflective equilibrium (Cohen 1981; Stein 1996; Stich 1990), and from evolution (Dennett 1987; Stich 1990) have famously claimed to have demonstrated uniformly optimal functioning of intentional-level psychologies in human beings and have been bolstered by the Apologist model that all normative/descriptive deviations can be explained in terms of computational limitations and performance errors.

The data on correlations with thinking dispositions provide one of two major ways of empirically addressing these claims. One is to determine the covariance among reasoning tasks after cognitive ability has been partialed out. The second is to examine whether there are cognitive/personality variables that can explain the normative/descriptive discrepancies that remain after computational limitations have been accounted for. Both of these methods as applied in our research program (and in those of some other commentators, **Epstein, Klaczynski, Newstead, Oaksford & Sellen**) have indicated that there is nonartifactual variation in intentional-level models of cognitive functioning. Again, granting **Funder**'s and **Krueger**'s point that using more psychometrically powerful measures of heuristics and biases will raise their correlations with cognitive ability (in cases where there is a positive correlation), it will also raise their correlations with measures of thinking dispositions. Thus, more powerful measures will not necessarily attenuate our initial finding that thinking dispositions do predict independent variance after cognitive ability has been controlled. Krueger is also right, however, in noting that improved measurement is likely to strengthen the argument that performance errors cannot explain norm violations.

R2. Rationality and dual process models

Friedrich calls our framing of the rationality issue within the dual process framework a bridge between the evolutionary and traditional biases perspectives, and this was indeed one of our goals – to create an integrated framework (**Adler** and **Over & Evans** also correctly see this as one of our goals). Contrary to **Hertwig**'s claim, we do not slight pragmatic intelligence. There is a section in the target article on the intelligence underlying pragmatic inferences and social intelligence (termed interactional intelligence in Table 3). **DeKay et al.,** like **Friedrich,** largely accept our characterization of evolutionary and normative rationality within the context of a generic two-process model. The broader evolutionary perspective that they call for is largely contained in Stanovich (1999) where there is a discussion that reiterates many of the points in their commentary. There is little in this commentary with which we disagree, with the exception of their characterization of Meliorism as claiming that humans are "fundamentally irrational" and are only "sometimes accidentally rational." We made it clear in the target article that in most cases the goals of Systems 1 and 2 will coincide and that System 1 processes will often *also* serve the goal of normative rationality (a point also stressed by Friedrich). Furthermore, in those minority of cases where a System 2 override is necessary in order for normative rationality (goal maximization for the individual) to be served, the operation of System 2 processes is hardly "accidental."

Acknowledging that the goals of the two systems often coincide goes some way toward recognizing the "complementarity" that **Frisch** wishes to stress. We do not deny that the two systems work in concert most of the time. But to us the most interesting cases (and those with potential real-world implications, see sects. 6.3 and R5) are those where the two systems compute conflicting information and cue different responses – one of which would violate a principle of normative rationality. As **Kahneman** points out, this is precisely the logic of the situation that captured the attention of the original heuristics and biases researchers and led to their seminal contributions. Nevertheless, we are not reticent to acknowledge situations such as those pointed out by **Frisch** and **Hardman,** where becoming more analytic can be detrimental. Stanovich (1999) contains a subsection titled "All System 1 overrides are not efficacious," where reference is made to the Wilson and Schooler (1991) study cited by **Hardman** in which they found situations in which a group of individuals who were encouraged to be analytic about their judgments made less effective decisions than a group who were not given encouragement to be analytic. This suggests that humans as cognitive systems can make the error of having too low a threshold for System 1 override. Such an overly low threshold for override might be conceived of as a Mr. Spock-like hyper-rationality that could actually be deleterious to goal achievement.

Funder, like **DeKay et al.,** views the original message of Meliorism as being that of "a fundamental shortcoming of the architecture of the human cognitive system causes its inferential processes inevitably to go awry" and feels that the newsworthiness of this message is undercut by the present conceptualization. But **Kahneman**'s commentary makes it clear that this was never the view of the original heuristics and biases researchers. Instead, their stance was in essence a two-process theory – as noted in the Kahneman commentary. Funder feels we have undercut the Meliorist message because, under our conception, inferential processes do not "inevitably go awry" since the goals of System 1 and 2 coincide in most cases. Second, even when they do not, there are individual differences in computational power and styles of epistemic regulation which determine System 2 override probability, and some individuals are thus able to compute the normatively correct response.

But the **Kahneman** commentary (and we believe the target article itself) makes clear why the Meliorist approach is not undercut by these findings. Although it is true that a particular normative *violation* will not be universal, the operation of the System 1 heuristic that could lead to such an error (if not overridden) *is* still universal within this conceptualization, and thus remains of great psychological interest. The computational biases inherent in this system are still ubiquitous and shared by all humans. There remains a newsworthy message in the Meliorist framework – and one which **Friedrich** (in his "bridge" comment) correctly views as a potential link between Meliorists and researchers in the cognitive ecologist camp. The two groups agree on certain aspects of the structure of System 1 processing and in some cases (e.g., Stanovich 1999) the adaptationist interpretation of System 1 processing. They disagree somewhat on the architecture of System 2 (see Samuels et al. 1999) and in the extent (and importance) of the situations in which System 1 and System 2 goals are in conflict (see sect. R5).

Numerous commentators made what are largely excellent suggestions for refining and supplementing the dual-process conceptualization in the target article. We have already commented on **Ball & Quayle**'s insightful discussion of computational escape hatches and the resolution of System 1 and System 2 outputs. **Moshman** (as, more indirectly, do **Stenning & Monaghan**) shares their emphasis on metacognition. We will mention several other useful notions in this section.

We would note, however, that to situate our contrast between normative and evolutionary rationality we set our findings within the context of a *generic* two-process model that addressed only those System 1/2 issues directly relevant to situating the notion of human rationality. We needed no more than a prototype model, one that did not choose between all of the micro-issues that might separate the theorists listed in our Table 3. If **Newstead** is arguing that there are not critical and fundamental family resemblances between these models, then we disagree with him. For example, Evans and Over (1996) explicitly discuss many direct parallels between their two-process model and that of Sloman (1996), and they rely heavily on Reber (1993) for their conceptualization of the tacit processing of System 1. Many other commonalities could be traced, but such an exercise is clearly beyond the scope of this Response.

More important, our concept of System 1/2 processing is not contradicted by the correlational evidence discussed by **Newstead.** Thinking dispositions are not associated with System 1 *or* System 2 as implied in that commentary. Because System 2 is the only system characterized by *flexible* goals (see sect. R2.1), it is variation in this system that the thinking disposition measures are assessing. System 1 and 2 are both whole systems – characterized by intentional level goal structures and algorithms that implement goal achievement. Styles are not associated with one system or the other. They are a (variable) property of the system (System 2) that employs epistemic and response regulation because its goals are flexible (see sect. R2.1 on goal flexibility). As discussed in section R1.7, psychometric intelligence relates to the System 2 algorithmic level, whereas thinking dispositions such as need for cognition and actively open-minded thinking relate to styles of epistemic and goal regulation at the intentional level of System 2. They are at different levels of analysis, and thus one would expect some degree of dissociation between the two. Intelligence is not synonymous with System 2 – it is a rough measure of the computational power available for System 2 overrides.

Newstead's summary of the state of the evidence on these issues also contrasts with how we view the literature. Klaczynski et al. (1997) have indeed found zero correlations between intelligence and degree of belief bias, but we have found small but significant correlations (Sá et al. 1999). More consistently, the meta-analysis by Cacioppo et al. (1996) indicates that there is a reliable correlation between need for cognition and intelligence, and in our own studies there are highly replicable correlations between intelligence and a variety of styles of epistemic regulation (Sá et al. 1999; Stanovich & West 1997).

We strongly agree with **Schneider, Stolarz-Fantino & Fantino,** and **Zizzo** that System 1 processes result from more than just innate specification. In our work on cognitive models of reading acquisition we were among the first to stress the importance of the concept of acquired modularity (Stanovich 1990; Stanovich et al. 1985) – in contrast to Fodor's (1983) concept. Thus, we agree that there is an important way in which the negative side effects of the fundamental computation biases in a technological economy can be overcome. Specifically, System 2 can strategically arrange practice so that habits of decontextualization become automatic.

Schneider is right to emphasize that automaticity is closely related to repetition and practice. Such a notion was stressed by the authors of two-process theories that predated the sample of more recent models listed in Table 3 (e.g., LaBerge & Samuels 1974; Posner & Snyder 1975; Shiffrin & Schneider 1977). For example, the Posner/Snyder two-process model is very similar to the alternative conceptualization laid out at the end of **Oberauer**'s commentary. Oberauer posits a spreading activation mechanism in memory just like Posner and Snyder (1975) and further posits that the nonselective activation of information by this system can be overcome when the reasoning system also has available sufficient working memory capacity and inhibitory mechanisms. The latter sound highly similar to the mechanisms in Posner and Snyder's (1975) conscious expectancy mechanism. All we need to add are Reber-like (1992a; 1992b; 1993) assumptions about the relative differences in evolutionary age of these two systems and we have a model not as different from our generic two-process synthesis as Oberauer implies.

We agree with **Schneider** and **Zizzo** that unconscious rule-based systems can be established through practice (this process would result in the automatic, rule-based cell of **Moshman**'s 2 x 2 partitioning), and thus we also agree that individual differences will sometimes arise in System 1 processes due to the different learning histories of the subjects as **Stolarz-Fantino & Fantino** and **Zizzo** argue. However, to the extent that these differences arise from conscious, System 2 decisions to structure experience and practice, some of these individual differences in System 1 will be parasitic on variance in System 2 strategic and metastrategic abilities.

Some errors can result from System 2 rules that have become inflexible because they have been instantiated as acquired System 1 algorithms. Something like this might be happening in the sunk cost research of Arkes and Ayton (1999) discussed in the **Ayton** commentary (and by **Jou**) – a general rule of "do not waste" might be triggering too automatically and unreflectively. Situations where acquired System 1 rules are overgeneralized may account for some of the reverse, and counterintuitive, developmental trends discussed by **Klaczynski** and **Reyna.**

The bottom-line, however, is that we are in sympathy with **Schneider**'s and **Zizzo**'s attempts to complicate our two-process dichotomy. We agree that aspects of the computational power and thinking dispositions of System 2 will create correlated variance in System 1 processes by means of the differential practice mentioned by these two commentators (and by **Stolarz-Fantino & Fantino**) – although how much variance this will create in System 1 processes is of course an empirical, and currently open, question.

We agree with **Friedrich** that in many situations System 2 will often act on the output of System 1 and thus the properties of System 1 will systematically infect System 2 processing. This was an aspect of Evans's original heuristic/analytic framework (Evans 1984; Evans & Wason 1976; Wason & Evans 1975; see the discussion in Evans & Over 1996), and we find it congenial because it accounts for instances of confabulation (Dennett 1991; Evans & Wason 1976; Gazzaniga 1998; Nisbett & Ross 1980; Nisbett & Wilson 1977) that are startlingly revealing of the modular structure of the brain.

Frisch, MacDonald & Geary, and **McCain** see more continuity across System 1 and 2 processing than we emphasize in the target article (a related point is made by **Newstead**). However, the evidence from neuropsychological dissociations and other evidence of process dissociations (Damasio 1994; Reber 1992a; 1992b; 1993; Sloman 1999; Sloman & Rips 1998; Smith et al. 1998; Willingham 1998; 1999) seems to be favoring the more separable dual-process model that we adopted in the target article, although the data on this issue are not yet definitive. Frisch, probably because of the abbreviated presentation in the target article, also misunderstands our model of why cognitive ability differences vary across types of selection tasks. It is described in detail in Stanovich and West (1998a) and in Chapter 4 of Stanovich (1999).

We agree with **Kahneman** that some people may make more nuanced System 1 judgments than others, and that individual differences in this capability are of some importance. This is related to **Teigen**'s point that when System 2 analytic abilities fail, well-framed intuitions may come to our assistance in narrowing the normative/descriptive gap, and the better those intuitions are the narrower the gap. But, following Reber (1992a; 1992b; 1993), we would conjecture that the variance in these System 1 abilities might well be considerably lower than the more recently evolved structures of System 2. Note, however, that this variability could become larger through the mechanism discussed above – instantiating of automatic System 1 algorithms through practice strategically initiated by System 2. Thus, some of the "well framed intuitions" referred to by Teigen may well be acquired intuitions – having their origins in capacity-intensive serial processing, yet now having the encapsulated, automatic characteristics of modular processes. Some statistics instructors, for example, become unable to empathize with their students for whom the basic probability axioms are not transparent. The instructor can no longer remember when these axioms were not primary intuitions.

Kahneman and **Teigen** are right to stress that System 1 processes support normatively rational behavior in many cases – as well as error (see also **Friedrich** and **MacDonald & Geary**). Regardless of the resolution of our speculation about the variability of System 1 processes, it is certainly clear that normative rationality is not served when aspects of System 1 functioning are *missing*. For example, most conceptions of emotions in cognitive science stress their adaptive regulatory powers. For example, in their discussion of the rationality of emotions, Johnson-Laird and Oatley (1992; see Oatley 1992) conceptualized emotions as interrupt signals supporting goal achievement. They see emotions as System 1 constructs that are particularly important in the characterization of systems whose behavior is governed by neither fixed action patterns nor impeccable rationality. Other cognitive scientists concur in this view (see Damasio 1994; de Sousa 1987). The basic idea is that emotions serve to stop the combinatorial explosion of possibilities that would occur if an intelligent system tried to calculate the utility of all possible future outcomes. Emotions are thought to constrain the possibilities to a manageable number based on somatic markers (see Damasio 1994) stored from similar situations in the past.

The work of Pollock (1991) is particularly relevant to the present discussion of the role of the emotions. In his view, heavily influenced by work in artificial intelligence, System 1 is composed of quick and inflexible (Q&I) modules that perform specific computations. In Pollock's (1995) model, emotions are conceived as Q&I modules for practical reasoning. Echoing the discussion of the "paranoid rabbits" in the **DeKay et al.** commentary, Pollock (1995) notes that "being afraid of tigers initiates quick avoidance responses without our having to think about it – a very useful reaction for anyone who is likely to encounter tigers unexpectedly. Embarrassment, indignation, and so forth, may similarly be practical Q&I modules whose purpose is to supplement explicit practical reasoning in social situations" (p. 11).

The key insight is that if we view emotions as Q&I modules for practical reasoning, there are two ways in which the rational regulation of behavior could go wrong. The two ways might be termed module failure and override failure, respectively. First, Q&I emotion modules might be missing or might malfunction. In this case, the automatic and rapid regulation of goals is absent, and System 2 is faced with a combinatorial explosion of possibilities because the constraining function of the emotions is missing. A module failure of this type represents a case where there is not too much emotion but instead too little.

The second way that behavioral regulation can go awry has the opposite properties. Here, the Q&I module has fired, but it happens to be one of those instances where the module's output is inappropriate and needs to be overridden by the controlled processing of System 2. Behavioral regulation is suboptimal when the System 2 override function does not work properly. It is clear that the folk psychological notion of the emotion/rationality relationship refers to the latter situation – failure to override System 1 Q&I modules for practical reasoning. This leads to the folk psychological cliché that emotion interferes with rational thought. But what folk psychology leaves out is irrationality of the first type. Here, the emotions play the opposite role – it is their absence that is the problem. Behavioral regulation fails to receive the crude but effective emotional signals that help to prioritize goals for subsequent action.

Important to note, there is empirical evidence for rationality failures of the two different types. Dorsolateral prefrontal damage has been associated with executive functioning difficulties (and/or working memory difficulties) that can be interpreted as the failure of System 2 to override automatized processes being executed by System 1 (Duncan et al. 1996; Kimberg et al. 1998; Kolb & Whishaw 1990; McCarthy & Warrington 1990; Shallice 1988). In contrast, ventromedial damage to the prefrontal cortex has been associated with problems in behavioral regulation that are accompanied by affective disruption (Bechara et al. 1994; Bechara et al. 1997; Damasio 1994). Difficulties of the former but not the latter kind are associated with lowered intelligence (Damasio 1994; Duncan et al. 1996) – consistent with the association of System 2 with psychometric intelligence and the relative independence of System 1 processes from the type of computational ability measured by IQ tests.

R2.1. The short-leash and long-leash goals of Systems 1 and 2. We are sympathetic to the suggestion of **Over & Evans** that a better term for normative rationality might be individual rationality (**Mandel** likewise critiques our use of the term normative rationality to describe instrumental rationality at the level of the individual). Much more important though, Over & Evans raise the issue of how we can assert that System 2 acts more to serve individual rationality than evolutionary rationality when it itself is an evolutionary product. Additional clarification is needed here because in their penultimate paragraph, **MacDonald & Geary** incorrectly view our discussion as mapping in an oversimplified way into the "levels of selection" issue in theoretical biology (Dawkins 1982; Hull 1992; Sober & Wilson 1994). Instead, our view follows from the interaction between the notion of long-leash and short-leash genetic control and what Dawkins (1976) calls the "essential tension" between the gene and its vehicle.

We conceptualize the differing goal structures of Systems 1 and 2 according to the "Mars Rover" metaphor used by Dennett (1984) and Plotkin (1988). Dennett (1984) describes how, when controlling a device such as a model airplane, one's sphere of control is only limited by the power of the equipment, but when the distances become large, the speed of light becomes a non-negligible factor. NASA engineers responsible for the Mars and Venus explorer vehicles knew that direct control was impossible because "the time required for a round trip signal was greater than the time available for appropriate action . . . Since controllers on Earth could no longer reach out and control them, they had to *control themselves*" (emphasis in original, p. 55). The NASA engineers had to move from the "short-leash" direct control, as in the model airplane case, to the "long-leash" control of the Mars explorer case where the vehicle is not given moment-by-moment instructions on how to act, but instead is given a more flexible type of intelligence plus some generic goals.

As Dawkins (1976) in his similar discussion of the Mars Rover logic in the science fiction story *A for Andromeda* notes, there is an analogy here to the type of control exerted by the genes when they build a brain:

> The genes can only do their best in *advance* by building a fast executive computer for themselves. . . . Like the chess pro-

> grammer, the genes have to "instruct" their survival machines not in specifics, but in the general strategies and tricks of the living trade. . . . The advantage of this sort of programming is that it greatly cuts down on the number of detailed rules that have to be built into the original program. (pp. 55, 57)

Human consciousness represents, according to Dawkins (1976)

> the culmination of an evolutionary trend towards the emancipation of survival machines as executive decision-makers from their ultimate masters, the genes. . . . By dictating the way survival machines and their nervous systems are built, genes exert ultimate power over behavior. But the moment-to-moment decisions about what to do next are taken by the nervous system. Genes are the primary policy-makers; brains are the executives. . . . The logical conclusion to this trend, not yet reached in any species, would be for the genes to give the survival machine a single overall policy instruction: do whatever you think best to keep us alive. (pp. 59–60)

This type of long-leash control that Dawkins is referring to is built on top of (that is, in *addition* to) the short-leash genetic control mechanisms that earlier evolutionary adaptation has installed in the brain. This relates to Dennett's (1991; 1996) four types of minds that are layered on top of each other. Dennett (1996) in his short but provocative book *Kinds of minds* describes the overlapping short-leashed and long-leashed strategies embodied in our brains by labelling them as different "minds" – all lodged within the same brain in the case of humans – and all simultaneously operating to solve problems.

One key distinction between Dennett's kinds of minds is how directly the various systems code for the goals of the genes. The Darwinian mind uses prewired reflexes and thus produces hardwired phenotypic behavioral patterns (the genes have "said" metaphorically "do *this* when x happens because it is best"). The Skinnerian mind uses operant conditioning to shape itself to an unpredictable environment (the genes have "said" metaphorically "learn what is best as you go along"). The Popperian mind can represent possibilities and test them internally before responding (the genes have "said" metaphorically "think about what is best before you do it"). The Gregorian mind (see Clark 1997) exploits the mental tools discovered by others (the genes have "said" metaphorically "imitate and use the mental tools used by others to solve problems").

In humans, all four "minds" are simultaneously operative. The Darwinian and Skinnerian minds – more akin to System 1 architectures – have short-leash goals installed ("when this stimulus appears, do *this*). These short-leash goals cause the Darwinian mind to always be ready to sacrifice the vehicle if it will aid genetic fitness. This is a point that **Greene & Levy** make in their commentary when they emphasize that genetic continuation is clearly more important than the optimal function of any individual within that species. In contrast, the Popperian and Gregorian minds have intentional-level psychologies characterized by long-leash goals ("operate with other agents in your environment so as to increase your longevity"). With the advent of the higher level minds, evolution has inserted into the architecture of the brain a flexible system that is somewhat like the ultimate long-leash goal suggested by Dawkins: "Do whatever you think best."

But "best for whom?" is the critical question here. The key is that for a creature with a flexible intelligence, long-leash goals, and a System 2 (Popperian/Gregorian) mind, we have the possibility of genetic optimization becoming dissociated from individual optimization. Consider the bees **Ayton** refers to in his commentary. As a hymenopteran (with their odd genetic relationships) a given bee will perform a number of acts in order to benefit its genetically related hive. As a creature with a Darwinian mind, an individual bee will even sacrifice itself as a vehicle if there is greater benefit to the same genes by helping other individuals (for instance, causing its own death when it loses its stinger while protecting its genetically-related hive-mates). There are no conflicting goals in a Darwinian creature. Its goals are the genes' goals pure and simple. Perfect rationality for the bee means local fitness optimization for its genes. Surely, as humans, we want more than that. In fact, a bee with Popperian intelligence and long-leash goals of self preservation might well decide that it could forgo the sacrifice! **Ayton** has made the same mistake that **Over & Evans** point out is sometimes made by evolutionary psychologists, which is to presuppose that what has evolutionary rationality also has individual rationality. We agree with Over & Evans that this is the source of serious confusion (and much unnecessary contention) in the field. To avoid the error, the different "interests" of the replicators and vehicles must be recognized – and we must keep definitions of rationality consistent with the entity whose optimization is at issue. To answer the question in Ayton's commentary title, the bee, as a Darwinian creature, needs no cognitive reform because it has no "interests" other than its genes' interests. Humans, with Gregorian minds, have interests as vehicles and thus might benefit from cognitive reform in situations where vehicle interests conflict with genetic interests and their Darwinian minds are siding with the latter.

In short, our conception is that at the intentional level, the goal structure of System 1 has been determined largely by evolutionary adaptation, whereas the goal structure of System 2 is more flexible and reflects ongoing goal evaluation at the personal level as an individual is shaped by environmental experience (see Stanovich 1999). Long-leash goals, and the System 2 mental structures necessary to satisfy them, lead to the separation between evolutionary and normative rationality discussed in the target article. To answer **Over & Evans,** this is how System 2 intelligence could be an adaptation yet side with individual rationality over evolutionary rationality in cases of conflict. To use a Dawkins-type (1976) phrasing, vehicles can rebel against the selfish replicators. That rebellion takes the form of optimizing vehicle utility rather than genetic fitness in cases where the two are in conflict.

When they started building Popperian and Gregorian minds, the genes were giving up on the strategy of coding moment-by-moment responses, and moving to a long-leash strategy that at some point was the equivalent of saying "Things will be changing too fast out there, brain, for us to tell you exactly what to do – you just go ahead and do what you think is best given the general goals (survival, sexual reproduction) that we (the genes) have inserted." And there is the rub. In long-leash brains, genetically coded goals can only be represented in the most general sense. There is no goal of "mate with female X at 6:57PM on Friday, June 13" but instead "have sex because it is pleasurable." But once the goal has become this general, a potential gap has been created whereby behaviors that might serve the vehicle's goal might not serve that of the genes. We need not go beyond the obvious example of sex with contraception, an act which serves the vehicle's goal of pleasure without serving the genes' goal of reproduction. What is happening here is that the flexible brain is coordinating multiple long-term goals – including its own survival and pleasure goals – and these multiple long-term goals come to overshadow its reproductive goal. From the standpoint of the genes, the human brain can sometimes be like a Mars Rover run amok. It is so busy coordinating its secondary goals (master your environment, engage in social relations with other agents, etc.) that it sometimes ignores the primary goal of replicating the genes that the secondary ones were supposed to serve.

Once we move beyond a Darwinian mind we begin to have parts of the brain that are devoted to gene replication in more indirect ways, through carrying out the general interests of the vehicle. Dawkins (1976) notes that there is an "uneasy tension . . . between gene and individual body as fundamental agent of life" (p. 234). **Ayton** has missed this essential tension by focusing on bees rather than humans. Darwinian creatures routinely sacrifice the vehicle in order to further the "interests" of the genes (see Dawkins, 1982, for numerous examples), which are of course replication. Only humans *really* turn the tables (or at least have the potential to) by occasionally ignoring the interests of the genes in order to further the interests of the vehicle.

Over & Evans present an interesting heritability puzzle in their commentary and resolve it by suggesting that highly intelligent people, by triggering greater cultural diffusion of knowledge (see Blackmore 1999), prevented their own level of System 2 ability from becoming widespread. Their suggestion is quite different from that of **MacDonald & Geary** who posit that System 2 structures are the direct result of selection pressures. Our own solution to the heritability puzzle mentioned by Over & Evans would be to imagine a bivariate space containing the scatterplot of two variables – the efficacy of a behavior in maximizing individual utility on the x-axis and the efficacy of a behavior in maximizing genetic fitness on the y-axis. Imagine a correlation of .80 – which would reflect the fact that genetic interests and vehicle interest most often coincide. If selection reduced the space to a thin bar at the top of and perpendicular to the y-axis (adaptation reducing variance and hence heritability as Over & Evans suggest) there would still be variance on the x-axis (although reduced in magnitude). Thus, a variant on Over & Evans's theme is to speculate that the heritability in intelligence that remains is the heritability in System 2 processes that are *specifically* related to fulfilling vehicle goals.

R2.2. The cultural evolution of normative standards. The interesting speculation of **Over & Evans** relates to a puzzling criticism made by **Schneider.** She seems to take the view that the cultural evolution of norms somehow presents difficulties for our conceptualization. In fact, we conjecture that a cultural history of norm evolution would support the generalization that individuals spawning progressive changes in rational standards are disproportionally of high intelligence and need for cognition. Braine and O'Brien (1991) argued exactly this when they stated

> judgments of logical validity may (implicitly) demand a distinction between inferences that depend only on the form of the information given and those that owe something to factual knowledge or pragmatic context (cf. Moshman & Franks, 1986). We speculate that the emergence of logic as a discipline required a level of civilization that included a subclass of intellectuals ready to put energy and persistence into the metacognitive (and metalogical) task. (p. 200)

But now through education the norms of rationality developed by that subclass with the energy and persistence for that metacognitive task are available to everyone. Our task, as learners, is much lightened of metacognitive load because these tools have already been constructed for us. This is essentially what Over & Evans argue.

In fact, contrary to **Schneider**'s implication, it is Panglossian theorists who have tended to ignore the implications of the historical evolution of normative standards. This comes about because of an emphasis on a competence/performance model – with its discrete emphasis on what principles are either in or not in reasoning competence – has impeded the appreciation of the fact that, particularly in the case of inductive inference, competence and experience are tightly intertwined (see Sophian 1997). The Panglossians (and the evolutionary psychologists) have tended to downplay the Gregorian mind, and instead are prone to extrapolate nativist assumptions from the domain of language into cognitive development more generally. This strong extrapolation of nativist assumptions (see Elman et al. 1996; Quartz & Sejnowski 1998) ignores the cultural history of norms and thus their potential learnability (see Krantz 1981). Inverting the figure and ground, Jepson et al. (1983) instead stress the historical contingency of the reasoning tools available to human reasoners: "the correctness of an induction depends not only on the adequacy of one's initial models but also on the conceptual tools one has available for extending or altering them. Changes in concepts and tools can therefore lead to different, and perhaps better, inductions. In other words, induction is a skill in which learning plays an important role" (Jepson et al. 1983, p. 495). Because normative models are tools of rationality for Gregorian minds to use, and because these tools undergo cultural change and revision, there is no idealized human "rational competence" that has remained fixed throughout history.

The evolution of reasoning norms has also been downplayed by Panglossians because of their tendency to leap to the defense of the majority of responders. But a change in standards could easily change a majority giving the "correct" response (from the Panglossian point of view) to the majority giving the "incorrect" one (which according to the Panglossian should never happen). That changing norms have this embarrassing implication is one reason why the historical evolution of norms is downplayed by these theorists. Equally embarrassing is the fact that majorities on some of the problems in the heuristics and biases literature can be changed with just a little experience, as demonstrated in Stanovich and West (1999). What invariably provokes the Panglossian critiques of heuristics and biases experiments is the finding that the modal subject departs from the normative response. The critiques are misguided if the finding that spawned them – that typical human performance departs from a specified normative model – displays systematic lability. Minimal prods to more thorough thinking can change the majority on some problems, leaving the Panglossian defending the minority instead of the majority.

R3. Process and rationality

Some commentators (e.g., **Hoffrage, Kahneman, Reyna**) point to the need for a process model of many of the tasks that were our focus. That is, these commentators wish to stress the necessity for a fully explicated model at the algorithmic level of analysis. We could not agree more that this is important and, in another domain of cognitive psychology, spent two decades doing just that (Stanovich 2000; Stanovich & West 1979; 1983; West & Stanovich 1978; 1986). However, this was not the purpose of the present research program. Rather than attempt a full unpacking of an algorithmic-level model for particular tasks, our purpose was to explore whether intentional-level models of human psychology must incorporate a notion of perfect rationality or whether, instead, such models should allow variation in that construct. To answer this question requires an exploration of the interplay between the intentional and algorithmic levels of analysis. An exclusive focus on the latter, while useful in its own right, simply could not answer the questions we wanted to address. We admit that individual differences are a very crude probe for the questions that we are interested in – but we submit that they are one of the few tools available for putting principled constraints on interpretations of the normative/descriptive gap. Thus, if any statements are to be made about the rational functioning of humans, then we need whatever tools we have, however crude.

In addition, we would point out that there is a venerable tradition behind the use of such intentional-level constructs by cognitive scientists – and for their use to be intermixed with algorithmic-level constructs in comprehensive theories. In a discussion of the intentional level of analysis, Newell (1982) argues that it allows us to "understand behavior without having an operational model of the processing that is actually being done by the agent" (p. 108). Prediction on this basis works "without requiring . . . the construction of any computational model" (p. 109). Newell (1982) further talks of enduring goal biases and epistemic criteria in a manner that places the thinking dispositions discussed in our target article clearly within the domain of intentional-level psychology: "The agent's goals must also be known . . . they are relatively stable characteristics that can be inferred from behavior and (for human adults) can sometimes by conveyed by language" (p. 108).

Newell (1982) explicitly discusses what he calls mixed systems of modeling and prediction, and argues that they are quite common in artificial intelligence as well as folk psychology, which both tend to mix processing notions from the algorithmic level with goal and belief notions from the intentional level. This is why we "recognize that forgetting is possible, and so we do not assume that knowledge once obtained is forever. We know that inferences are only available if the person thinks it through, so we don't assume that knowing X means knowing all the remote consequences of X, though we have no good way of determining exactly what inferences will be known" (p. 115). Newell (1982) argues that such mixed models are often better than models based on only one level of conceptual analysis.

To the extent that the intentional level of analysis often implicates issues of rationality (Dennett 1987; Samuels et al. 1999; Sloman 1999), then our research program might be said to be a type of empirical philosophy – and of course this risks raising the hackles of both philosophers and empirical psychologists alike. However, several commentators (e.g., **Oaksford & Sellen, Over & Evans**) are sympathetic to our program because, we feel, they thoroughly appreciate that a focus on intentional-level constructs (rationality, dispositions) does not detract at all from the quest for a fuller algorithmic-level specification. In fact, a venerable tradition in cognitive science (Anderson 1990; 1991; Levelt 1995; Marr 1982; Newell 1982; Oaksford & Chater 1995) supports the notion that there can be synergistic interplay between levels. In a nuanced article, Levelt (1995) argues that, if anything, psychology has been afflicted by "processitis" – an overly exclusive emphasis on algorithmic-level models that has left analyses of human intentional states to the disciplines of linguistics, sociology, anthropology, and economics. Indeed, one could view the interdisciplinary field of cognitive science as reflecting an attempt to integrate sciences focused on the algorithmic level of analysis (e.g., psychology) with sciences focused on the intentional level (e.g., anthropology, economics) in order to more fully understand mental functioning. Related to Newell's (1982) comment quoted above, Levelt (1995) argues that the very struggle to correctly assign an explanatory factor to one level or another can be informative for cognitive science.

R4. Performance errors

Bucciarelli, Hoffrage, and **Hunt** prefer a different gloss on the distinction between performance errors and computational limitations than the one we used. They would treat many instances of what we term computational limitations as performance errors. While we admit that there are weaknesses within our own conception of performance errors (without care, our conception can be made to seem a strawman), the parsing favored by these commentators suffers from problems that are equally severe. Hoffrage is clearest in his divergence with our view in calling recurring motivational and attentional problems performance errors. There are several problems with parsing the distinction in this manner. Stable motivational errors are in fact like cognitive styles and thinking dispositions at the intentional level – they reflect stable and predictable behavioral tendencies. If stable and predictable, these behavioral tendencies should be conceptualized as variation in modes of epistemic and behavioral regulation, as variation in rationality at the intentional level. Such generality in modes of behavioral regulation is exactly what Baron (1985b) invokes when defining the notion of a cognitive style, an intentional-level construct in our view. Similarly, recurring attentional problems reflect, in our view, the type of stability that is best viewed as a computational limitation. Certainly this is how most investigators view the recurring attentional and executive disorders that characterize individuals with attention-deficit/hyperactivity disorder or dorsolateral frontal lobe damage (Duncan et al. 1996; Kimberg et al. 1998; Kolb & Whishaw 1990; McCarthy & Warrington 1990; Pennington & Ozonoff 1996; Shallice 1988).

Finally, the parsing of the performance error/computational limitation distinction in the manner of these commentators is inconsistent with the spirit of the Panglossian position in both psychology and economics. When the performance error argument is invoked by Cohen (1981) to refute the heuristics and biases researchers, or invoked by neoclassical economists in defense of their rationality assumption, the clear implication is that such errors should be considered trivial in the context of overall human performance. In the latter case, for example, performance errors will be arbitraged away by the majority of other (rational) market participants and optimal equilibria will be restored. But arbitrage opportunities that are *repeatedly missed* are another matter. They do not have the trivial implications assumed when the performance error notion is invoked. They undermine the assumption of optimality that is the foundation of the neoclassical analysis. In short, a performance error that keeps repeating is not a performance error at all – as indicated in the quotes from Frank (1990) and Thaler (1992) in section 2 of the target article. Such recurring errors are inconsistent with the Panglossian perspective, which views them as an absolving explanation.

R5. The fundamental computational bias and "real life"

Some commentators (e.g., **Funder, Hardman, Jou**) take us to task for a focus on so-called unrepresentative problem situations: situations which subjects do not face on a day-to-day basis. These commentators question our contention that, ironically, life is becoming more like the tests! Other commentators, however, (e.g., **Baron, DeKay et al., Friedrich, Kahneman, Klaczynski, Over & Evans**) concurred with our emphasis on the importance of situations where the modern world presents evolutionarily adapted mechanisms with problems they were not designed to solve. We find nothing in the commentaries to alter our view that the fundamental computational bias (demonstrated in both laboratory and nonlaboratory research) can have important negative consequences when it leads to non-normative responses by individuals faced with a real-world task requiring cognitive decontextualization. And we continue to maintain that requirements for cognitive decontextualization are increasing in modern society (as shown in the analyses of Gottfredson 1997). These increasing requirements create more opportunities for System 1/2 mismatches where normative rationality is not aligned with the evolutionary rationality that characterizes System 1 processes.

For example, it used to be the case that people ate only what was produced in their immediate environments. Now, when they go to the supermarket, a panoply of foods from all over the world is presented to them, and their choices are made in the context of an advertising-saturated society that presents messages that they wish to avoid but cannot (Wilson & Brekke 1994).

Technicians of modern mass communication have become quite skilled at implying certain conclusions without actually stating those conclusions (for fear of lawsuits, bad publicity, etc.). Advertisements rely on the fundamental computational bias (particularly its enthymematic processing feature) to fill in the missing information. The communication logic of relying on System 1 processes to trump System 2 is readily employed by advertisers, in election campaigns, and even by governments – for example in promoting their lottery systems ("You could be the one!" blares an ad from the Ontario Lottery Commission, thereby increasing the availability of an outcome which, in the game called 6/49, has an objective probability of 1 in 14 million).

Thus, in order to achieve one's goals in a technological society where normative rationality and evolutionary rationality have come apart, the evolutionarily adaptive responses of System 1 will increasingly have to be overridden by the strategic, capacity demanding operations of System 2. In fact, such dissociations are a major theme in the writings of the cognitive ecologists. Cosmides and Tooby (1996) argue that

> in the modern world, we are awash in numerically expressed statistical information. But our hominid ancestors did not have access to the modern accumulation which has produced, for the first time in human history, reliable, numerically expressed statistical information about the world beyond individual experience. Reliable numerical statements about single event probabilities were rare or nonexistent in the Pleistocene. (p. 15)
>
> It is easy to forget that our hominid ancestors did not have access to the modern system of socially organized data collection, error checking, and information accumulation. . . . In ancestral environments, the only external database available from which to reason inductively was one's own observations. (Brase et al. 1998, p. 5)

Precisely. We are living in a technological society where we must: decide which health-care provider to join based on just such statistics; figure out whether to invest in a retirement annuity; decide what type of mortgage to purchase; figure out what type of deductible to get on our auto insurance; decide whether to trade in our car or sell it ourselves; decide whether to lease or to buy; think about how to apportion our TIAA/CREF retirement funds; and decide whether we would save money by joining a book club. And we must make all of these decisions based on information represented in a manner for which our brains are not adapted (in none of these cases have we coded frequency information from our own personal experience). In order to reason normatively in all of these domains (in order to maximize our own personal utility) we are going to have to deal with probabilistic information represented in nonfrequentistic terms – in representations that the cognitive ecologists have shown are different from our well-adapted algorithms for dealing with frequency information (Cosmides & Tooby 1996; Gigerenzer & Hoffrage 1995).

Several commentators such as **Ayton** and **Hardman** emphasized how efficient the "fast and frugal" strategies studied by Gigerenzer and Goldstein (1996; Gigerenzer, Todd, and the ABC Research Group 1999) can be in certain situations. These System 1 processes are no doubt extremely efficacious in certain domains but, as **Kahneman** notes, all such heuristics introduce weighting biases, biases that we need explicit System 2 strategies and knowledge to correct. There is nothing wrong with the current emphasis on such fast and frugal heuristics, but to stress them *exclusively* carries over the unhelpful tendency from evolutionary psychology to ignore the contribution of the Gregorian mind (Clark 1997; Dennett 1991; 1995; 1996) – the mind that is uniquely human (slow and expensive though it may be). The explicit normative rules stored there as tools of cultural evolution are sometimes needed in order to override the weighting biases of System 1 heuristics.

Consider the work of Brase et al. (1998), who improved performance on the notorious three-card problem (Bar-Hillel & Falk 1982; Falk 1992; Granberg 1995) by presenting the information as frequencies and in terms of whole objects, both alterations designed to better fit the frequency-computation systems of the brain. In response to a query about why the adequate performance observed was not even higher given that our brains contain such well-designed frequency-computation systems, Brase et al. (1998) replied that "in our view it is remarkable that they work on paper-and-pencil problems at all. A natural sampling system is designed to operate on actual events" (p. 13). The problem is that in a symbol-oriented postindustrial society, we are presented with paper-and-pencil problems all the time, and much of what we know about the world comes not from the perception of actual events but from abstract information preprocessed, prepackaged, and condensed into symbolic codes such as probabilities, percentages, tables, and graphs (the voluminous statistical information routinely presented in *USA Today* comes to mind).

We have here an example of the figure and ground reversals that permeate contemporary rationality studies and that were well captured in a chapter by Samuels et al. (1999) which we encountered after preparing our target article. They stated that

> we suspect that those Panglossian-inclined theorists who describe Darwinian modules as "elegant machines" are tacitly assuming that normative evaluation should be relativized to the proper domain, while those who offer a bleaker assessment of human rationality are tacitly relativizing their evaluations to the actual domain, which, in the modern world, contains a vast array of information-processing challenges that are quite different from anything our Pleistocene ancestors had to confront. (p. 114)

Thus, it is possible to accept most of the conclusions of the evolutionary psychologists but to draw completely different morals from them. The evolutionary psychologists want to celebrate the astonishing job that evolution did in adapting to that Pleistocene environment. Certainly they are right to do so. The more we understand about evolutionary mechanisms, the more awed appreciation we have for them. But at the same time, it is not inconsistent for a thoroughgoing Meliorist to be horrified at the fact that a multi-million dollar advertising industry is in part predicated on creating stimuli that will trigger System 1 heuristics that many of us will not have the cognitive energy or cognitive disposition to override. For a Meliorist, it is no consolation that the heuristics so triggered were evolutionarily adaptive in their day.

Neither is it a consolation to be told that such situations are not typical. That is no help when the very few situations where System 1/2 mismatches are likely to occur have enormous knock-on effects. **Friedrich** may well be right that, on a purely quantitative basis, it may appear that System 2 activity is being overestimated in the research literature. In terms of the micro-events in day-to-day life, that is no doubt true. Throughout the day we are detecting frequencies hundreds of times, detecting faces dozens of times, using our language modules repeatedly, inferring intentionality in other human beings constantly, and so on. But the very few instances where we require System 2 may be of unusual importance. The modern consumer world is littered with System 1 traps and, often, the more potentially costly the situation, the more such traps there are (automobile purchases, mutual fund investments, mortgage closing costs, and insurance come to mind).

We would reiterate that although reinforcement learning will help, as **Zizzo** argues, it is no panacea for the problems that technological society poses for a cognitive system characterized by the fundamental computational bias. The act of buying baked beans (and small standard commodities like them – see **Zizzo**'s discussion) occurs on dozens of occasions that provide opportunities for just the type of learning that economists have demonstrated (see Zizzo's citations), but the complex contingencies and tradeoffs present in choosing a mortgage option or in signing an employment contract may occur only a few times in a lifetime. For these decisions, there is every indication that we need the serial, capacity-demanding, contingency calculation of System 2. Finally, to answer **Frisch**'s closing comment, we do indeed think that it is a valid statistical generalization to say that for the majority of people in the first world, regrettably, "life" is being lived more and more in the "human made world." To point out this historical trend (which is likely to continue until population stabilizes) is not to endorse it. Note that understanding the point in the last sentence requires just the type of detachment and decontextualization that we discussed in section 6.3.

Schneider and other commentators (e.g., **Funder, Hardman**) ask for evidence that, to paraphrase, "more intelligent people enjoy more rational outcomes in the real world." We in fact cited such evidence in the target article. The psychometric literature contains numerous indications that cognitive ability is correlated with the avoidance of harmful behaviors and with success in employment settings independent of level of education (Brody 1997; Gottfredson 1997; Hunt 1995; Lubinski & Humphreys 1997) as well as social status attainment as noted by **MacDonald & Geary.** In short, the very types of outcomes that normative rationality should maximize for individuals are reasonably strongly predicted by the analytic abilities that comprise System 2.

R6. Potentially productive ideas

Many commentaries contained interesting ideas that can only be mentioned here. **Stenning & Monaghan** unpack our System 1/2 properties in a somewhat different but interesting way. They discuss two dimensions – cooperative versus adversarial communication and explicit versus implicit knowledge – and describe how they reparse the System 1/2 distinction.

Moshman elaborates System 1 and System 2 differently. He makes the interesting suggestion that a more complete conceptualization is obtained by crossing automatic versus explicit processing with heuristic versus rule-based processing. It is of course more customary in traditional two-process views to conjoin *what* the system computes (e.g., rules) with *how* the processing is implemented (e.g., by a virtual von Neumann serial processor simulated by a connectionaist network; see Dennett 1991). Moshman suggests that certain developmental trends are handled with greater explanatory power if we dissociate these two things.

Okasha is quite right to note that not all instances of reject-the-norm arguments stem from Panglossian biases (**Frisch** makes the same point). Some authors reject the norm for independent reasons – not simply because of a Panglossian reflex to close all normative/descriptive gaps.

Indeed, we cited an example of this in the target article. Dawes (1990) – hardly a Panglossian theorist – argued (rightly we believe) that the wrong normative model has been applied to the opinion-prediction task that defines the so-called false consensus effect. Note that in some sense it is the Meliorist camp that is freer to reject the norm for independent reasons because, unlike the Panglossians, they are not forced to defend the majority in every instance. Also, note that the commitment to the naturalistic fallacy is much stronger among Panglossians than among Meliorists. The normative models and task construals of the Panglossians are *determined* by the majority responders. In contrast, our commitment to the naturalistic fallacy is considerably weaker. It is only that empirical results regarding individual differences represent *one* (of many) type of information that might be considered in a wide reflective equilibrium of experts (Daniels 1979; 1996; Sloman 1999; Stanovich 1999; Stein 1996; Stich & Nisbett 1980). If **Ayton** does not like the choice of the normative model for a given situation to be determined by a vote, then his real beef is with the Panglossian camp.

Kahneman is certainly right that there are various notions of rationality throughout the sprawling philosophical and empirical literature on the topic (Audi 1993; Cook & Levi 1990; Elster 1983; Foley 1987; Gibbard 1990; Harman 1995; Nathanson 1994; Nozick 1993). The reasoning-rationality versus coherence-rationality distinction described in his commentary cross-cuts several other dichotomies such as instrumental versus theoretical rationality and the rationality$_1$ and rationality$_2$ of **Evans & Over** (1996; see also, Sloman 1999). Although it might not be apparent in the target article, our empirical program encompasses more aspects of rationality than simply reasoning rationality. Stanovich (1999) describes studies of knowledge calibration and overconfidence in which coherence rationality is the main issue. We also have examined between-subjects analyses of several framing effects: the type of analyses that Kahneman suggests isolate a test of coherence rationality. We also note that we *did* use the eight-item Linda version that Kahneman classifies as more similar to a between-subjects design than the within-subjects design, and thus is closer to a coherence rationality assessment. Finally, we strongly agree with **Kahneman** that most of life resembles a between-subjects design, and thus our sampling of within-subjects situations if anything biased a good portion of our research program in the Panglossian direction.

Reyna and **Klaczynski** allude to the possibility of applying the understanding/acceptance principle using developmental data. This is a suggestion that we are sympathetic with, as noted in Note 5 of the target article (see also **Moshman**'s commentary). There are complications in such an application however. Often, in such studies, it is necessary to make fundamental changes in the task in order to span different ages, and this makes clean developmental comparisons difficult. This accounts for some of the reverse developmental trends (non-normative responses increasing with age) cited by **Reyna** and **Klaczynski** and termed "paradoxical" by **Hertwig.** The use of this term by Hertwig is puzzling because he resolves the seeming paradox at the end of his own commentary in exactly the way that we would – in class inclusion tasks the term "probability" is used with adults and permits many more interpretations than does the term "more" that is used with children. As a result, it is artifactual that children's performance can be made to seem higher. There is no paradox here at all. Similarly, with the base-rate studies by Davidson (1995) and Jacobs and Potenza (1991) cited by Hertwig, Klaczynski, and Reyna – the paradox disappears when the details of the studies are examined. The diagnosticity of the indicant information in these studies is dependent on knowledge of a stereotype that increases with age (Billy likes dancing and is thus more likely to prefer cooking to football). Thus, younger children may seem to be using base-rate information more because the indicant information is unavailable to them since they lack the stereotype. Note that use of the base-rate does not decrease with age in the "object" condition of Jacobs and Potenza (1991), as opposed to the "social" condition, because in the object condition the indicant information is not dependent on knowledge of a stereotype. Many of the "paradoxical" developmental trends in the heuristics and biases literature can be explained in a similar manner.

Many more excellent suggestions abound in these commentaries. **Girotto** illustrates how variations in non-optimal choice patterns may be relevant for testing rational choice assumptions and how such variation can sometimes lead to successful collective actions. **Bucciarelli**'s suggestions for examining trial-to-trial strategic variation and for building non-deterministic features into reasoning theories are well taken. **Reyna** offers fuzzy trace theory as an alternative two-process account to the one we present. **Stolarz-Fantino & Fantino**'s emphasis on looking to subjects' histories of decision making for explanations of processing variance is a good one. **Wang** describes an interesting study showing the importance of the social structure of task environments.

Of course we agree with **Hoffrage** and **Okasha** that a finer-grained scoring system is needed to fully explicate performance in the cabs and AIDS problems – and that the labels on the categories we used (Bayesian, indicant, etc.) are not to be taken literally. In the original analyses of individual differences on these problems, we attempted many more fine-grained partitionings. But for the broad-based purposes of the target article, alternative partitionings do not change the essential conclusions. Finally, we have evidence supporting **Kahneman**'s point about difficult problems like the AIDS problem – that if the problem were made easier, more talented respondents would be most likely to benefit from improved problem features, and thus with these improved versions the correlation between the use of base rates and intelligence should turn positive. Stanovich and West (1999) produced just this pattern by removing the confusing term "false positive rate" from the problem.

R7. Complementary strengths and weaknesses

Frisch calls for a treatment of the so-called rationality debate that reflects more complementarity among perspectives. This is not difficult to do. It is perhaps best accomplished by recognizing the unique contributions of each of the camps and how they fit together – for complementarity is easily discernible when one views the strengths and the weaknesses of each of the "camps" in the context of each other (see Stanovich, 1999, for a more extended discussion). For example, each perspective has advanced the theoretical and practical debate about human rationality: the Panglossians have demonstrated that the answers to real-world problems are often in human cognition itself (economists have long exploited this strategy by assuming that humans are already optimizing in the real-world and that we simply need to properly characterize the optimized processes); Apologists (the cognitive ecologists) have demonstrated the power of evolutionary explanations and the necessity of matching stimulus representations to those to which evolution has shaped cognition (e.g., Cosmides & Tooby 1996); the Meliorists have championed the possibility of cognitive change (e.g., Nisbett 1993) and warned against the dire consequences of mismatches between human cognitive tendencies and the thinking requirements of a technological society (e.g., Piattelli-Palmarini 1994).

Correspondingly, however: the Meliorists are often too quick to claim flaws in reasoning and they may miss real strengths in human cognition that could be exploited for behavioral betterment; the Apologists sometimes fail to acknowledge that a real cognitive disability results when a technological society confronts the human cognitive apparatus with a representation for which it is not evolutionarily adapted – and they sometimes fail to stress that representational flexibility is something that can increase with instruction; and the Panglossians are often forced into contorted positions in order to excuse human errors, and they thus pass up opportunities to remediate correctable cognitive mistakes.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

Abelson, R. P., Leddo, J. & Gross, P. H. (1987) The strength of conjunctive explanations. *Personality and Social Psychology Bulletin* 13:141–55. [SS-F]

Adler, J. E. (1984) Abstraction is uncooperative. *Journal for the Theory of Social Behaviour* 14:165–81. [aKES]

(1991) An optimist's pessimism: Conversation and conjunctions. In: *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science,* ed. E. Eells & T. Maruszewski. Editions Rodopi. [aKES]

(1994) Fallacies and alternative interpretations. *Australasian Journal of Philosophy* 74:271–82. [JEA]

Ainslie, G. (1992) *Picoeconomics.* Cambridge University Press. [rKES]

Ajzen, I. (1977) Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology* 35:303–14. [aKES]

Alker, H. & Hermann, M. (1971) Are Bayesian decisions artificially intelligent? The effect of task and personality on conservatism in information processing. *Journal of Personality and Social Psychology* 19:31–41. [aKES]

Allais, M. (1953) Le comportement de l'homme rationnel devant le risque: Critique des postulate et axioms de l'ecole americaine. *Econometrica* 21:503–46. [PA, DF, arKES]

Alloy, L. B. & Tabachnik, N. (1984) Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review* 91:112–49. [aKES]

Anderson, J. R. (1990) *The adaptive character of thought.* Erlbaum. [arKES]

(1991) Is human cognition adaptive? *Behavioral and Brain Sciences* 14:471–517. [JJ, arKES]

Anderson, J. R. & Lebiere, C. (1998) *The atomic components of thought.* Erlbaum. [KO]

Anderson, N. H. (1981) *Foundations of information integration theory.* Academic Press. [SS-F]

Arkes, H. R. & Ayton, P. (1999) The sunk cost and concorde effects: Are humans less rational than lower animals? *Psychological Bulletin* 125:591–600. [PA, JJ, rKES]

Arkes, H. R. & Blumer, C. (1985) The psychology of sunk cost. *Organizational Behavior and Human Decision Processes* 35:124–40. [DH]

Arkes, H. R. & Hammond, K., eds. (1986) *Judgment and decision making.* Cambridge University Press. [aKES]

Audi, R. (1993) *Action, intention, and reason.* Cornell University Press. [rKES]

Axelrod, R. (1984) *The evolution of cooperation.* Basic Books. [VG]

Ayton, P. & Arkes, H. (1998) Call it quits. *New Scientist* 158:40–43. [PA]

Ayton, P. & Hardman, D. (1997) Are two rationalities better than one? *Current Psychology of Cognition* 16:39–51. [aKES]

Baddeley, A. (1994) Working memory: The interface between memory and cognition. In: *Memory systems 1994,* ed. D. L. Schacter & E. Tulving. Bradford/MIT Press. [KM]

Ball, L. J. & Quayle, J. D. (1999) The effects of belief and logic on syllogistic reasoning: Evidence from an eye-tracking analysis. In: *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society,* ed. M. Hahn & S. C. Stones. Erlbaum. [LJB]

Ball, L. J., Sutherland, E. J. & Quayle, J. D. (1997) Situating deductive competence within a models framework: Some further considerations. *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology – Current Psychology of Cognition* 16:52–62. [LJB]

Bara, B. G., Bucciarelli, M. & Johnson-Laird, P. N. (1995) The development of syllogistic reasoning. *American Journal of Psychology* 108:157–93. [MB, arKES]

Bara, B., Bucciarelli, M. & Lombardo, V. (2000) Model theory of deduction: A unified computational approach. *Cognitive Science.* (in press). [MB]

Bargh, J. A. & Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist* 54:462–79. [JF]

Bar-Hillel, M. (1980) The base-rate fallacy in probability judgments. *Acta Psychologica* 44:211–33. [aKES]

(1990) Back to base rates. In: *Insights into decision making: A tribute to Hillel J. Einhorn,* R. M. Hogarth. University of Chicago Press. [aKES]

(1991) Commentary on Wolford, Taylor, and Beck: The conjunction fallacy? *Memory and Cognition* 19:412–14. [rKES]

Bar-Hillel, M. & Falk, R. (1982) Some teasers concerning conditional probabilities. *Cognition* 11:109–22. [rKES]

Barkow, J. H. (1989) *Darwin, sex, and status: Biological approaches to mind and culture.* University of Toronto Press. [aKES]

Baron, J. (1985a) *Rationality and intelligence.* Cambridge University Press. [arKES]

(1985b) What kinds of intelligence components are fundamental? In: *Thinking and learning, vol. 2,* ed. S. Chipman & J. Segal. Erlbaum. [rKES]

(1993) *Morality and rational choice.* Kluwer. [arKES]

(1994) Nonconsequentialist decisions. *Behavioral and Brain Sciences* 17:1–42. [arKES]

(1995) My-side bias in thinking about abortion. *Thinking and Reasoning* 1:221–35. [aKES]

(1998) *Judgment misguided: Intuition and error in public decision making.* Oxford University Press. [arKES]

(in press) *Thinking and deciding, 3rd edition.* Cambridge University Press. [JB]

Baron, J. & Hershey, J. C. (1988) Outcome bias in decision evaluation. *Journal of Personality and Social Psychology* 54:569–79. [aKES]

Barrouillet, P. & Lecas, J. (1999) Mental models in conditional reasoning and working memory. *Thinking and Reasoning* 5:289–302. [rKES]

Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50:7–15. [rKES]

Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–95. [rKES]

Bell, D., Raiffa, H. & Tversky, A., eds. (1988a) *Decision making: Descriptive, normative, and prescriptive interactions.* Cambridge University Press. [aKES]

(1988b) Descriptive, normative, and prescriptive interactions in decision making. In: *Decision making: Descriptive, normative, and prescriptive interactions,* ed. D. Bell, H. Raiffa & A. Tversky. Cambridge University Press. [DRM]

Berkeley, D. & Humphreys, P. (1982) Structuring decision problems and the "bias heuristic." *Acta Psychologica* 50:201–52. [aKES]

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

Blackmore, S. (1999) *The meme machine.* Oxford University Press. [rKES]

Block, J. & Funder, D. C. (1986) Social roles and social perception: Individual differences in attribution and "error." *Journal of Personality and Social Psychology* 51:1200–207. [DCF, JK, aKES]

Boulding, K. E. (1956) *The image.* University of Michigan Press. [RAM]

Braine, M. D. S. & O'Brien, D. P. (1991) A theory of "if": A lexical entry, reasoning program and standard logic. *Psychological Review* 98:182–203. [rKES]

Braine, M. D. S. & O'Brien, D. P., eds. (1998) *Mental logic.* Erlbaum. [DM]

Brainerd, C. J. (1981) Working memory and the developmental analysis of probability judgment. *Psychological Review* 88:463–502. [PAK]

Brainerd, C. J. & Reyna, V. F. (1990) Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review* 10:365–403. [VFR]

(1992a) Explaining "memory free" reasoning. *Psychological Science* 3:332–39. [VFR]

(1992b) The memory independence effect: What do the data show? What do the theories claim? *Developmental Review* 12:164–86. [VFR]

(1993) Memory independence and memory interference in cognitive development. *Psychological Review* 100:42–67. [VFR]

Brainerd, C. J., Reyna, V. F. & Mojardin, A. H. (1999) Conjoint recognition. *Psychological Review* 106:160–79. [VFR]

Brase, G. L., Cosmides, L. & Tooby, J. (1998) Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General* 127:3–21. [arKES]

Bratman, M. E., Israel, D. J. & Pollack, M. E. (1991) Plans and resource-bounded practical reasoning. In: *Philosophy and AI: Essays at the interface,* ed. J. Cummins & J. Pollock. MIT Press. [arKES]

Brody, N. (1997) Intelligence, schooling, and society. *American Psychologist* 52:1046–50. [arKES]

Broome, J. (1990) Should a rational agent maximize expected utility? In: *The limits of rationality,* ed. K. S. Cook & M. Levi. University of Chicago Press. [arKES]

Bruner, J. (1986) *Actual minds, possible worlds.* Harvard University Press. [aKES]

(1990) *Acts of meaning.* Harvard University Press. [aKES]

Bucciarelli, M. & Johnson-Laird, P. N. (1999) Strategies in syllogistic reasoning. *Cognitive Science* 23:247–303. [MB, VG, MO, aKES]

Byrne, R. M. J. (1989) Suppressing valid inferences with conditionals. *Cognition* 31:1–21. [MO]

Byrnes, J. P. (1998) *The nature and development of decision-making.* Erlbaum. [VFR]

Byrnes, J. P. & Overton, W. F. (1986) Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology* 22:793–99. [aKES]

Cacioppo, J. T., Petty, R. E., Feinstein, J. & Jarvis, W. (1996) Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin* 119:197–253. [JF, arKES]

Camerer, C. & Ho, T. H. (1999) Experience-weighted attraction learning in normal form games. *Econometrica* 67:827–74. [DJZ]

Cameron, N. (1939) Schizophrenic thinking in a problem-solving situation. *Journal of Mental Science* 85:1012–35. [MO]

(1954) Experimental analysis of schizophrenic thinking. In: *Language and thought in schizophrenia,* ed. J. S. Kasanin. University of California Press. [MO]

Campos, J. J., Campos, R. G. & Barrett, K. C. (1989) Emergent themes in the study of emotional development and emotion regulation. *Developmental Psychology* 25:394–402. [KM]

Capra, F. (1982) *The turning point: Science, society and the rising culture.* Simon & Schuster. [DF]

Caporael, L. R. (1997) The evolution of truly social cognition: The core configuration model. *Personality and Social Psychology Review* 1:276–98. [KM]

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* 97:404–31. [KM, aKES]

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

(1997) Psychometrics, intelligence, and public perception. *Intelligence* 24:25–52. [aKES]

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

Case, D. A., Fantino, E. & Goodie, A. S. (1999) Base-rate training without case cues reduces base-rate neglect. *Psychonomic Bulletin and Review* 6:319–27. [SS-F]

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

Ceci, S. J. (1996) *On intelligence: A bioecological treatise on intellectual development* (expanded edition). Harvard University Press. [arKES]

Ceci, S. J. & Williams, S. J. (1997) Schooling, intelligence, and income. *American Psychologist* 52:1051–58. [rKES]

Chaiken, S. & Trope, Y. (1999) *Dual process theories in social psychology.* Guilford. [DK]

Chapman, L. J. & Chapman, J. P. (1973) *Disordered thought in schizophrenia.* Appleton-Century-Crofts. [MO]

Chater, N. & Oaksford, M. (1994) A rational analysis of the selection task as optimal data selection. *Psychological Review* 101:608–31. [KS]

Chen, E. Y. H., Wilkins, A. J. & McKenna, P. J. (1994) Semantic memory is both impaired and anamolous in schizophrenia. *Psychological Medicine* 24:193–202. [MO]

Cheng, P. W. & Holyoak, K. J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology* 17:391–416. [KIM]

(1989) On the natural selection of reasoning theories. *Cognition* 33:285–313. [aKES]

Cheng, P. W. & Novick, L. R. (1992) Covariation in natural causal induction. *Psychological Review* 99:365–82. [UH]

Cherniak, C. (1986) *Minimal rationality.* MIT Press. [aKES]

Chi, M. T. H., Glaser, R. & Farr, M. J., eds. (1988) *The nature of expertise.* Erlbaum. [SLS]

Clark, A. (1997) *Being there: Putting brain, body, and world together again.* MIT Press. [rKES]

Clark, H. & Clark, E. (1979) *The psychology of language.* Harcourt Brace Jovanovich. [VFR]

Clemen, R. T. (1999) Does decision analysis work? A research agenda. Duke University. (unpublished manuscript). [DH]

Cohen, L. J. (1979) On the psychology of prediction: Whose is the fallacy? *Cognition* 7:385–407. [aKES]

(1981) Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences* 4:37–70. [DH, arKES]

(1982) Are people programmed to commit fallacies? Further thoughts about the interpretation of experimental data on probability judgment. *Journal for the Theory of Social Behavior* 12:251–74. [arKES]

(1983) The controversy about irrationality. *Behavioral and Brain Sciences* 6:510–17. [aKES]

(1986) *The dialogue of reason.* Oxford University Press. [aKES]

Cook, K. S. & Levi, M., eds. (1990) *The limits of rationality.* University of Chicago Press. [rKES]

Cooper, W. S. (1989) How evolutionary biology challenges the classical theory of rational choice. *Biology and Philosophy* 4:457–81. [aKES]

Cooper, W. S. & Kaplan, R. H. (1982) Adaptive "coin-flipping": A decision-theoretic examination of natural selection for random individual variation. *Journal of Theoretical Biology* 94:135–51. [AJG]

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

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

(1996) Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58:1–73. [arKES]

Cummins, D. D. (1996) Evidence for the innateness of deontic reasoning. *Mind and Language* 11:160–90. [aKES]

Cummins, D. D., Lubart, T., Alksnis, O. & Rist, R. (1991) Conditional reasoning and causation. *Memory and Cognition* 19:274–82. [MO]

Cunningham, A. E. & Stanovich, K. E. (1997) Early reading acquisition and its relation to reading experience and ability ten years later. *Developmental Psychology* 33:934–45. [rKES]

Damasio, A. R. (1994) *Descartes' error.* Putnam. [rKES]

Daniels, N. (1979) Wide reflective equilibrium and theory acceptance ethics. *The Journal of Philosophy* 76:256–82. [rKES]

(1996) *Justice and justification: Reflective equilibrium in theory and practice.* Cambridge University Press. [rKES]

Darwin, C. (1859/1959) *On the origin of species by means of natural selection. A variorum text,* ed. M. Pickham. (Reprinted 1959). University of Pennsylvania Press. [AJG]

Daston, L. (1980) Probabilistic expectation and rationality in classical probability theory. *Historia Mathematica* 7:234–60. [aKES]

(1988) *Classical probability in the enlightenment.* Princeton University Press. [PA]

Davidson, D. (1995) The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill-Palmer Quarterly* 41:328–46. [RH, PAK, VFR, rKES]

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

(1990) The potential nonfalsity of the false consensus effect. In: *Insights into decision making,* ed. R. M. Hogarth. University of Chicago Press. [arKES]

(1998) Behavioral decision making. In: *Handbook of social psychology* (4th edition), ed. D. T. Gilbert, S. T. Fiske & G. Lindzey. McGraw-Hill. [JK]

Dawkins, R. (1976) *The selfish gene* (New edition, 1989). Oxford University Press. [arKES]

(1982) *The extended phenotype.* Oxford University Press. [arKES]

Deary, I. J. (1995) Auditory inspection time and intelligence: What is the direction of causation? *Developmental Psychology* 31:237–50. [aKES]

Deary, I. J. & Stough, C. (1996) Intelligence and inspection time. *American Psychologist* 51:599–608. [aKES]

De Finetti, B. (1970) *Theory of probability, vol. 1.* Wiley. (republished 1990). [aKES]

De Long, J. B., Shleifer, A., Summers, L. H. & Waldmann, R. J. (1990) Noise trader risk in financial markets. *Journal of Political Economy* 98:703–38. [DJZ]

Dempster, F. (1992) The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review* 12:45–75. [VFR]

Denes-Raj, V. & Epstein, S. (1994) Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology* 66(5):819–29. [DH]

Dennett, D. C. (1980) The milk of human intentionality. *Behavioral and Brain Sciences* 3:428–30. [aKES]

(1984) *Elbow room: The varieties of free will worth wanting.* MIT Press. [rKES]

(1987) *The intentional stance.* MIT Press. [arKES]

(1988) Precis of *The intentional stance. Behavioral and Brain Sciences* 11:493–544. [aKES]

(1991) *Consciousness explained.* Little, Brown. [rKES]

(1995) *Darwin's dangerous idea: Evolution and the meanings of life.* Simon & Schuster. [rKES]

(1996) *Kinds of minds: Toward an understanding of consciousness.* Basic Books. [rKES]

Desmond, N. L. & Levy, W. B. (1997) Ovarian steroidal control of connectivity in the female hippocampus: An overview of recent experimental finding and speculations on its functional consequence. *Hippocampus* 7:239–45. [AJG]

De Sousa, R. (1987) *The rationality of emotion.* MIT Press. [rKES]

Detterman, D. K. (1994) Intelligence and the brain. In: *The neuropsychology of individual differences,* ed. P. A. Vernon. Academic Press. [aKES]

Dinsmoor, J. A. (1985) The role of observing and attention in establishing stimulus control. *Journal of the Experimental Analysis of Behavior* 43:365–81. [AJG]

Doherty, M. E. & Mynatt, C. (1990) Inattention to P(H) and to P(D/-H): A converging operation. *Acta Psychologica* 75:1–11. [aKES]

Doherty, M. E., Schiavo, M., Tweney, R. & Mynatt, C. (1981) The influence of feedback and diagnostic data on pseudodiagnosticity. *Bulletin of the Psychonomic Society* 18:191–94. [aKES]

Dominowski, R. L. (1995) Content effects in Wason's selection task. In: *Perspectives on thinking and reasoning,* ed. S. E. Newstead & J. St. B. T. Evans. Erlbaum. [aKES]

Donaldson, M. (1978) *Children's minds.* Fontana. [aKES]

(1993) *Human minds: An exploration.* Viking Penguin. [aKES]

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

Duncan, J., Burgess, P. & Emslie, H. (1995) Fluid intelligence after frontal lobe lesions. *Neuropsychologia* 33:261–68. [rKES]

Duncan, J., Emslie, H., Williams, P., Johnson, R. & Freer, C. (1996) Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology* 30:257–303. [rKES]

Economist, The (1998) *The benevolence of self-interest.* December 12, 1998, p. 80. [aKES]

Edwards, W. (1992) Toward the demise of economic man and woman: Bottom lines from Santa Cruz. In: *Utility theories: Measurements and applications,* ed. W. Edwards. Kluwer. [PA]

Einhorn, H. J. (1982) Learning from experience and suboptimal rules in decision making. In: *Judgment under uncertainty: Heuristics and biases,* ed. D. Kahneman, P. Slovic & A. Tversky. Cambridge University Press. [JJ]

Einhorn, H. J. & Hogarth, R. M. (1981) Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology* 32:53–88. [aKES]

Ellsberg, D. (1961) Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75:643–99. [JB, DF]

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking innateness: A connectionist perspective on development.* MIT Press. [rKES]

Elster, J. (1988) *Sour grapes: Studies in the subversion of rationality.* Cambridge University Press. [arKES]

(1989) *The cement of society.* Cambridge University Press. [VG]

Engle, R. W., Tuholski, S. W., Laughlin, J. E. & Conway, A. R. A. (1999) Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General* 128:309–31. [rKES]

Ennis, R. H. (1987) A taxonomy of critical thinking dispositions and abilities. In: *Teaching thinking skills: Theory and practice,* ed. J. Baron & R. Sternberg. W. H. Freeman. [rKES]

Epstein, S. (1973) The self-concept revisited or a theory of a theory. *American Psychologist* 28:404–16. [SE]

(1994) Integration of the cognitive and the psychodynamic unconscious. *American Psychologist* 49:709–24. [LJB, SE, DF, aKES]

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* 62:328–39. [aKES]

Epstein, S. & Pacini, R. (1999) Some basic issues regarding dual-process theories from the perspective of cognitive-experiential self-theory. In: *Dual process theories in social psychology,* ed. S. Chaiken & Y. Trope. Guilford. [SE]

Epstein, S., Pacini, R., Denes-Raj, V. & Heier, H. (1996) Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology* 71:390–405. [SEN]

Erev, I. & Roth, A. E. (1998) Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* 88(4):848–81. [RAM]

Estes, W. K. & Straugham, J. H. (1954) Analysis of verbal conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology* 47:225–34. [AJG]

Evans, J. St. B. T. (1982) *The psychology of deductive reasoning.* Routledge. [aKES]

(1984) Heuristic and analytic processes in reasoning. *British Journal of Psychology* 75:451–68. [arKES]

(1989) *Bias in human reasoning: Causes and consequences.* Erlbaum. [KO, aKES]

(1996) Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology* 87:223–40. [aKES]

(2000) What could and could not be a strategy in reasoning? In: *Deductive reasoning and strategies,* ed. W. Schaeken, G. De Vooght, A. Vandierendonck & G. d'Ydewalle. Erlbaum. [SEN]

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

Evans, J. St. B. T. & Lynch, J. S. (1973) Matching bias in the selection task. *British Journal of Psychology* 64:391–97. [aKES]

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

Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and reasoning.* Psychology Press. [LJB, DEO, arKES]

(1997) Rationality in reasoning: The problem of deductive competence. *Current Psychology of Cognition* 16:3–38. [DEO]

Evans, J. St. B. T., Over, D. E. & Manktelow, K. (1993) Reasoning, decision making and rationality. *Cognition* 49:165–87. [LJB, rKES]

Evans, J. St. B. T. & Wason, P. C. (1976) Rationalization in a reasoning task. *British Journal of Psychology* 67:479–86. [rKES]

Falk, R. (1992) A closer look at the probabilities of the notorious three prisoners. *Cognition* 43:197–223. [rKES]

Fantino, E., Kulik, J., Stolarz-Fantino, S. & Wright, W. (1997) The conjunction fallacy: A test of averaging hypotheses. *Psychonomic Bulletin and Review* 4:96–101. [SS-F]

Fantino, E. & Savastano, H. I. (1996) Humans' responses to novel stimulus compounds and the effect of training. *Psychonomic Bulletin and Review* 3:204–207. [DJZ]

Fiedler, K. (1988) The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research* 50:123–29. [aKES]

Fischbacher, U. & Falk, R. (1999) Kindness is the parent of kindness – Modeling reciprocity. Paper presented at the Allied Social Science Associations Convention, New York, 1999. [RAM]

Fischhoff, B. (1996) The real world: What good is it? *Organizational Behavior and Human Decision Processes* 65(3):232–48. [DH]

Fishburn, P. C. (1991) Nontransitive preferences in decision-theory. *Journal of Risk and Uncertainty* 4:113–34. [PA]

Fodor, J. (1983) *Modularity of mind.* MIT Press. [rKES]

Foley, R. (1987) *The theory of epistemic rationality.* Harvard University Press. [rKES]

Fong, G. T., Krantz, D. H. & Nisbett, R. E. (1986) The effects of statistical training on thinking about everyday problems. *Cognitive Psychology* 18:253–92. [aKES]

Ford, M. (1994) Two modes of mental representation and problem solving in syllogistic reasoning. *Cognition* 54:1–71. [MO]

Frank, R. H. (1990) Rethinking rational choice. In: *Beyond the marketplace,* ed. R. Friedland & A. Robertson. Aldine de Gruyter. [arKES]

Frank, R. H. & Cook, P. J. (1995) *The winner-take-all society.* The Free Press. [aKES]

Frank, R. H., Gilovich, T. & Regan, T. (1993) Does studying economics inhibit cooperation? *Journal of Economic Perspectives* 7:159–71. [DJZ]

Friedrich, J. (1987) Perceived control and decision making in a job hunting context. *Basic and Applied Social Psychology* 8:163–76. [JF]

(1993) Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review* 100:298–319. [JF, aKES]

Frisch, D. (1993) Reasons for framing effects. *Organizational Behavior and Human Decision Processes* 54:399–429. [aKES]

(1994) Consequentialism and utility theory. *Behavioral and Brain Sciences* 17:16. [aKES]

Frisch, D. & Jones, S. (1993) Assessing the accuracy of decisions. *Theory and Psychology* 3:115–35. [DF]

Fry, A. F. & Hale, S. (1996) Processing speed, working memory, and fluid intelligence. *Psychological Science* 7:237–41. [aKES]

Funder, D. C. (1987) Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin* 101:75–90. [MO, arKES]

Galef, B. G. (1998) Recent progress in studies of imitation and social learning in animals. In: *Advances in psychological science, vol. 2: Biological and cognitive aspects,* ed. M. Sabourin. Psychology Press/Erlbaum. [AJG]

Galotti, K. M. (1995) A longitudinal study of real-life decision making: Choosing a college. *Applied Cognitive Psychology* 9:459–84. [DH]

Gardner, H. (1983) *Frames of mind: The theory of multiple intelligence.* Basic. [RJS]

(1999) Are there additional intelligences? The case for naturalist, spiritual, and existential intelligence. In: *Education, information, and transformation,* ed. J. Kane. Prentice-Hall. [RJS]

Gauthier, D. (1986) *Morals by agreement.* Oxford University Press. [rKES]

Gavanski, I. & Roskos-Ewoldson, D. R. (1991) Representativeness and conjoint probability. *Journal of Personality and Social Psychology* 61:181–94. [SS-F]

Gazzaniga, M. S. (1998) *The mind's past.* University of California Press. [rKES]

Geary, D. C. (1998) *Male, female: The evolution of human sex differences.* American Psychological Association. [KM]

Gibbard, A. (1990) *Wise choices, apt feelings: A theory of normative judgment.* Harvard University Press. [rKES]

Gigerenzer, G. (1991a) From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review* 98:254–67. [aKES]

(1991b) How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology* 2:83–115. [DF, ASG, aKES]

(1993) The bounded rationality of probabilistic mental models. In: *Rationality: Psychological and philosophical perspectives,* ed. K. Manktelow & D. Over. Routledge. [aKES]

(1994) Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In: *Subjective probability,* ed. G. Wright & P. Ayton. Wiley. [JJ]

(1996a) On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review* 103:592–96. [DF, aKES]

(1996b) Rationality: Why social context matters. In: *Interactive minds: Life-span perspectives on the social foundation of cognition,* ed. P. B. Baltes & U. Staudinger. Cambridge University Press. [PA, aKES]

Gigerenzer, G. & Goldstein, D. G. (1996) Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103(4):650–69. [PA, DH, arKES]

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

Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991) Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review* 98:506–28. [UH, aKES]

Gigerenzer, G. & Hug, K. (1992) Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition* 43:127–171. [KIM]

Gigerenzer, G. & Regier, T. (1996) How do we tell an association from a rule? Comment on Sloman (1996). *Psychological Bulletin* 119:23–26. [RH, aKES]

Gigerenzer, G., Todd, P. M. & The ABC Research Group (1999) *Simple heuristics that make us smart.* Oxford University Press. [ASG, DH, DK, rKES]

Gilovich, T. (1991) *How we know what isn't so: The fallibility of human reason in everyday life.* The Free Press. [JEA]

Girotto, V. & Gonzalez, M. (2000) Strategies and models in statistical reasoning. In: *Strategies in deductive reasoning,* ed. W. Schaeken, G. De Vooght, A. Vandierendonck & G. d'Ydewalle. Erlbaum. [VG]

Goldman, A. I. (1978) Epistemics: The regulative theory of cognition. *Journal of Philosophy* 55:509–23. [aKES]

Goodie, A. S. & Fantino, E. (1996) Learning to commit or avoid the base-rate error. *Nature* 380:247–49. [SS-F, DJZ]

(1999) What does and does not alleviate base-rate neglect under direct experience. *Journal of Behavioral Decision Making* 12:307–35. [ASG]

Gottfredson, L. S. (1997) Why g matters: The complexity of everyday life. *Intelligence* 24:79–132. [KM, arKES]

Granberg, D. (1995) The Monte Hall dilemma. *Personality and Social Psychology Bulletin* 31:711–23. [rKES]

Greenfield, P. M. (1997) You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist* 52:1115–24. [aKES]

Grice, P. (1975) Logic and conversation. In: *Syntax and semantics, vol. 3: Speech acts,* ed. P. Cole & J. Morgan. Academic Press. [KS]

(1978) Further notes on logic and conversation. In: *Syntax and semantics,* ed. P. Cole. Academic Press. [VFR]

(1989) *Studies in the way of words.* Harvard University Press. [JEA]

Griggs, R. A. (1983) The role of problem content in the selection task and in the THOG problem. In: *Thinking and reasoning: Psychological approaches,* ed. J. St. B. T. Evans. Routledge & Kegan Paul. [aKES]

Griggs, R. A. & Cox, J. R. (1982) The elusive thematic-materials effects in Wason's selection task. *British Journal of Psychology* 73:407–20. [aKES]

(1983) The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology* 35:519–33. [aKES]

Hammond, K. R. (1996) *Human judgment and social policy.* Oxford University Press. [DF, aKES]

Handley, S. J., Newstead, S. E. & Wright, H. (2000) Rational and experiential thinking: A study of the REI. In: *International perspectives on individual differences: Vol. 1. Cognitive styles,* ed. R. J. Riding & S. G. Rayner. [SEN]

Hardman, D. K. & Payne, S. J. (1995) Problem difficulty and response format in syllogistic reasoning. *Quarterly Journal of Experimental Psychology* 48A:945–75. [LJB]

Harman, G. (1995) Rationality. In: *Thinking, vol. 3,* ed. E. E. Smith & D. N. Osherson. MIT Press. [arKES]

Hartl, J. A. & Fantino, E. (1996) Choice as a function of reinforcement ratios in delayed matching to sample. *Journal of the Experimental Analysis of Behavior* 66:11–27. [SS-F]

Haselton, M. G. & Buss, D. M. (in press) Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology.* [WTD]

Haselton, M. G., Buss, D. M. & DeKay, W. T. (1998) A theory of errors in cross-sex mind reading. Paper presented at the Human Behavior and Evolution Society Conference, Tucson, AZ, July 1998. [WTD]

Haslam, N. & Baron, J. (1994) Intelligence, personality, and prudence. In: *Personality and intelligence,* ed. R. J. Sternberg & P. Ruzgis. Cambridge University Press. [rKES]

Hawkins, J., Pea, R. D., Glick, J. & Scribner, S. (1984) "Merds that laugh don't like mushrooms": Evidence for deductive reasoning by preschoolers. *Developmental Psychology* 20:584–94. [DM]

Henle, M. (1962) On the relation between logic and thinking. *Psychological Review* 69:366–78. [arKES]

(1978) Foreword. In: *Human reasoning,* ed. R. Revlin & R. Mayer. Wiley. [arKES]

Herrnstein, R. J. & Murray, C. (1994) *The bell curve.* Free Press. [RJS]

Hertwig, R. & Chase, V. M. (1998) Many reasons or just one: How response mode affects reasoning in the conjunction problem. *Thinking and Reasoning* 4:319–52. [RH]

Hertwig, R. & Gigerenzer, G. (1999) The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making* 12:275–305. [PA, RH, KO]

Hilton, D. J. (1995) The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin* 118:248–71. [RH, DRM, aKES]

Hoch, S. J. (1987) Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology* 53:221–34. [aKES]

Hoch, S. J. & Tschirgi, J. E. (1985) Logical knowledge and cue redundancy in deductive reasoning. *Memory and Cognition* 13:453–62. [aKES]

Hoffrage, U. & Gigerenzer, G. (1999) How to foster diagnostic inferences in experts. (submitted). [UH]

Howson, C. & Urbach, P. (1993) *Scientific reasoning: The Bayesian approach* (second edition). Open Court. [aKES]

Huber, B. L. & Huber, O. (1987) Development of the concept of comparative subjective probability. *Journal of Experimental Child Psychology* 44:304–16. [DM]

Hull, D. L. (1982) The naked meme. In: *Learning, development, and culture: Essays in evolutionary epistemology,* ed. H. C. Plotkin. Wiley. [aKES]

(1992) A matter of individuality. In: *The units of selection: Essays on the nature of species,* ed. M. Ereshefsky. MIT Press. [rKES]

Hunt, E. (1980) Intelligence as an information-processing concept. *British Journal of Psychology* 71(4):449–74. [EH]

(1987) The next word on verbal ability. In: *Speed of information-processing and intelligence,* ed. P. A. Vernon. Ablex. [aKES]

(1995) *Will we be smart enough? A cognitive analysis of the coming workforce.* Russell Sage Foundation. [arKES]

(1997) Nature vs. nurture: The feeling of vuja de. In: *Intelligence, heredity, and environment,* ed. R. J. Sternberg & E. L. Grigorenko. Cambridge University Press. [aKES]

(1999) Intelligence and human resources: Past, present, and future. In: *Learning and individual differences: Process, trait, and content determinants,* ed. P. Ackerman, P. Kyllonen & R. Richards. American Psychological Association. [rKES]

Inhelder, B. & Piaget, J. (1964) *The early growth of logic in the child.* Norton. [RH]

Jacobs, J. E. & Potenza, M. (1991) The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development* 62:166–78. [PAK, VFR, arKES]

Jensen, A. R. (1998) *The g factor. The science of mental ability.* Praeger/Greenwood. [EH, KM, RJS]

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

Johnson-Laird, P. N. (1991) *Human and machine thinking.* Erlbaum. [MB]

(1983) *Mental models: Towards a cognitive science of language, inference and consciousness.* Harvard University Press. [MB, aKES]

(1995) Inference and mental models. In: *Perspectives on thinking and reasoning: Essays in honour of Peter Wason,* ed. S. E. Newstead & J. St. B. T. Evans. Erlbaum. [LJB]

(1999) Deductive reasoning. *Annual Review of Psychology* 50:109–35. [aKES]

Johnson-Laird, P. N. & Byrne, R. M. J. (1990) Meta-logical puzzles: Knights, knaves and Rips. *Cognition* 36:69–84. [VG]

(1991) *Deduction.* Erlbaum. [LJB, MB, aKES]

(1993) Models and deductive rationality. In: *Rationality: Psychological and philosophical perspectives,* ed. K. Manktelow & D. Over. Routledge. [VG, aKES]

Johnson-Laird, P. N. & Oatley, K. (1992) Basic emotions, rationality, and folk theory. *Cognition and Emotion* 6:201–23. [arKES]

Johnson-Laird, P. N. & Steedman, M. (1978) The psychology of syllogisms. *Cognitive Psychology* 10:64–99. [MB]

Jones, K. & Day, J. D. (1997) Discrimination of two aspects of cognitive-social intelligence from academic intelligence. *Journal of Educational Psychology* 89:486–97. [aKES]

Jou, J. & Shanteau, J. (1995) Gestalt and dynamic processes in decision making. *Behavioral Processes* 33:305–18. [JJ]

Jou, J., Shanteau, J. & Harris, R. J. (1996) An information processing view of framing effects: The role of causal schemas in decision making. *Memory and Cognition* 24:1–15. [aKES]

Jungermann, H. (1986) The two camps on rationality. In: *Judgment and decision making,* ed. H. R. Arkes & K. R. Hammond. Cambridge University Press. [aKES]

Kahneman, D. (1981) Who shall be the arbiter of our intuitions? *Behavioral and Brain Sciences* 4:339–40. [aKES, XTW]

Kahneman, D., Slovic, P. & Tversky, A., eds. (1982) *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press. [aKES]

Kahneman, D. & Tversky, A. (1982) On the study of statistical intuitions. *Cognition* 11:123–41. [aKES] Also in: *Judgment under uncertainty: Heuristics and biases,* ed. D. Kahneman, P. Slovic & A. Tversky. Cambridge University Press. [DK]

(1983) Can irrationality be intelligently discussed? *Behavioral and Brain Sciences* 6:509–10. [aKES]

(1984) Choices, values, and frames. *American Psychologist* 89:341–50. [DF, aKES]

(1996) On the reality of cognitive illusions. *Psychological Review* 103:582–91. [RH, DK, arKES]

Kardash, C. M. & Scholes, R. J. (1996) Effects of pre-existing beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology* 88:260–71. [arKES]

Keating, D. P. & Hertzman, C., eds. (1999) *Developmental health and the wealth of nations: Social, biological, and educational dynamics.* Guilford Press. [rKES]

Kimberg, D. Y., D'Esposito, M. & Farah, M. J. (1998) Cognitive functions in the prefrontal cortex - working memory and executive control. *Current Direction in Psychological Science* 6:185–92. [rKES]

Klaczynski, P. A. & Fauth, J. (1997) Developmental differences in memory-based intrusions and self-serving statistical reasoning biases. *Merrill-Palmer Quarterly* 43:539–66. [VFR]

Klaczynski, P. A. & Gordon, D. H. (1996) Goal-directed everyday problem solving: Motivational and general ability influences on adolescent statistical reasoning. *Child Development* 67:2873–91. [PAK]

Klaczynski, P. A., Gordon, D. H. & Fauth, J. (1997) Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology* 89:470–85. [SEN, arKES]

Klaczynski, P. A. & Narasimhan, G. (1998) Problem representation as mediators of adolescent deductive reasoning. *Developmental Psychology* 34:865–81. [PAK]

Klaczynski, P. A. & Robinson, B. (in press) Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning biases. *Psychology and Aging.* [PAK]

Klahr, D., Fay, A. L. & Dunbar, K. (1993) Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology* 25:111–46. [aKES]

Klar, Y. & Giladi, E. E. (1997) No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology* 73:885–901. [JK]

Klayman, J. & Ha, Y. (1987) Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review* 94:211–28. [aKES]

Klein, G. (1998) *Sources of power: How people make decisions.* MIT Press. [DF, DH, aKES]

Kleindorfer, P. R., Kunreuther, H. C. & Schoemaker, P. J. H. (1993) *Decision sciences: An integrative perspective.* Cambridge University Press. [DRM]

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

Kolb, B. & Whishaw, I. (1990) *Fundamentals of human neuropsychology* (third edition). W. H. Freeman. [rKES]

Koriath, A., Lichtenstein, S. & Fischhoff, B. (1980) Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory* 6:107–18. [aKES]

Kornblith, H. (1993) *Inductive inference and its natural ground.* MIT Press. [aKES]

Kornblith, H., ed. (1985) *Naturalizing epistemology.* MIT Press. [aKES]

Krantz, D. H. (1981) Improvements in human reasoning and an error in L. J. Cohen's. *Behavioral and Brain Sciences* 4:340–41. [arKES]

Kreps, D. M. & Wilson, R. (1982) Reputation and imperfect information. *Journal of Economic Theory* 27:253–79. [DJZ]

Krueger, J. (1998a) The bet on bias: A foregone conclusion? *Psycoloquy* 9(46). http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?9.46 [JK]

(1998b) On the perception of social consensus. *Advances in Experimental Social Psychology* 30:163–240. [JK]

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. [aKES]

Krueger, J. & Zieger, J. (1993) Social categorization and the truly false consensus effect. *Journal of Personality and Social Psychology* 65:670–80. [aKES]

Kruglanski, A. W. & Webster, D. M. (1996) Motivated closing the mind: "Seizing" and "freezing." *Psychological Review* 103:263–83. [rKES]

Kuhberger, A. (1995) The framing of decisions: A new look at old problems. *Organizational Behavior and Human Decision Processes* 62:230–40. [aKES]

Kuhn, D. (2000) Metacognitive development. In: *Child psychology: A handbook of contemporary issues,* ed. L. Balter & C. S. Tamis-LeMonda. Psychology Press. [DM]

Kyburg, H. E. (1983) Rational belief. *Behavioral and Brain Sciences* 6:231–73. [aKES]

(1991) Normative and descriptive ideals. In: *Philosophy and AI: Essays at the interface,* ed. J. Cummins & J. Pollock. MIT Press. [aKES]

Kyllonen, P. C. (1996) Is working memory capacity Spearman's *g*? In: *Human abilities: Their nature and measurement,* ed. I. Dennis & P. Tapsfield. Erlbaum. [arKES]

Kyllonen, P. C. & Christal, R. E. (1990) Reasoning ability is (little more than) working memory capacity?! *Intelligence* 14:389–433. [arKES]

LaBerge, D. & Samuels, S. (1974) Toward a theory of automatic information processing in reading. *Cognitive Psychology* 6:293–323. [rKES]

Larrick, R. P., Morgan, J. N. & Nisbett, R. E. (1990) Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science* 1:362–70. [DH]

Larrick, R. P., Nisbett, R. E. & Morgan, J. N. (1993) Who uses the cost-benefit rules of choice? Implications for the normative status of microeconomic theory. *Organizational Behavior and Human Decision Processes* 56:331–47. [SLS, arKES]

Larrick, R. P., Smith, E. E. & Yates, J. F. (1992) Reflecting on the reflection effect: Disrupting the effects of framing through thought. Paper presented at the meetings of the Society for Judgment and Decision Making, St. Louis, MO, November 1992. [aKES]

Leibniz, G. W. (1677/1951) Towards a universal characteristic. In: *Leibniz: Selections,* ed. P. Wiener. Scribner's. [PA]

Levelt, W. (1995) Chapters of psychology. In: *The science of the mind: 2001 and beyond,* ed. R. L. Solso & D. W. Massaro. Oxford University Press. [rKES]

Levi, I. (1981) Should Bayesians sometimes neglect base rates? *Behavioral and Brain Sciences* 4:342–43. [SO]

(1983) Who commits the base rate fallacy? *Behavioral and Brain Sciences* 6:502–506. [SO, aKES]

Levinson, S. C. (1995) Interactional biases in human thinking. In: *Social intelligence and interaction,* ed. E. Goody. Cambridge University Press. [aKES]

Lichtenstein, S., Fischhoff, B. & Phillips, L. (1982) Calibration and probabilities: The state of the art to 1980. In: *Judgment under uncertainty: Heuristics and biases,* ed. D. Kahnemann, P. Slovic & A. Tversky. Cambridge University Press. [aKES]

Lichtenstein, S. & Slovic, P. (1971) Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89:46–55. [aKES]

Lopes, L. L. (1981a) Performing competently. *Behavioral and Brain Sciences* 4:343–44. [aKES]

(1981b) Decision making in the short run. *Journal of Experimental Psychology: Human Learning and Memory* 7:377–85. [DF]

(1982) Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:626–36. [aKES]

(1984) Risk and distributional inequality. *Journal of Experimental Psychology: Human Perception and Performance* 10:465–85. [RJS]

(1987) Between hope and fear: The psychology of risk. *Advances in Experimental Social Psychology* 20:255–95. [XTW]

(1988) Economics as psychology: A cognitive assay of the French and American schools of risk theory. In: *Risk, decision and rationality,* ed. B. R. Munier. D. Reidel. [rKES]

(1990) Re-modeling risk aversion: A comparison of Bernollian and rank dependent value approaches. In: *Acting under uncertainty: Multidisciplinary conceptions,* ed. G. M. von Fuerstenberg. Kluwer. [XTW]

(1991) The rhetoric of irrationality. *Theory and Psychology* 1:65–82. [aKES]

(1996) When time is of the essence: Averaging, aspiration, and the short run. *Organizational Behavior and Human Decision Processes* 65:179–89. [DF]

Lopes, L. L. & Oden, G. C. (1991) The rationality of intelligence. In: *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science,* ed. E. Eells & T. Maruszewski. Editions Rodopi. [arKES]

Lubinski, D. & Humphreys, L. G. (1997) Incorporating general intelligence into epidemiology and the social sciences. *Intelligence* 24:159–201. [arKES]

Luce, R. D. (1991) Rank- and sign-dependent linear utility representations. *Journal of Economic Theory* 53:75–100. [XTW]

(1992) Where does subjective expected utility fail descriptively? *Journal of Risk and Uncertainty* 5:5–27. [XTW]

Luria, A. R. (1976) *Cognitive development: Its cultural and social foundations.* Harvard University Press. [aKES]

Lykken, D. T. & Tellegen, A. (1993) Is human mating adventitious or the result of lawful choice? A twin study of mate selection. *Journal of Personality and Social Psychology* 65:56–68. [AJG]

Lynn, R. (1997) *Dysgenics.* Praeger. [KM]

Lyon, D. & Slovic, P. (1976) Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica* 40:287–98. [aKES]

Macchi, L. (1995) Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental Psychology* 48A:188–207. [aKES]

MacCrimmon, K. R. (1968) Descriptive and normative implications of the decision- theory postulates. In: *Risk and uncertainty,* ed. K. Borch & J. Mossin. Macmillan. [aKES]

MacCrimmon, K. R. & Larsson, S. (1979) Utility theory: Axioms versus 'paradoxes.' In: *Expected utility hypotheses and the Allais paradox,* ed. M. Allais & O. Hagen. D. Reidel. [aKES]

MacDonald, K. B. (1991) A perspective on Darwinian psychology: The importance of domain-general mechanisms, plasticity, and individual differences. *Ethology and Sociobiology* 12:449–80. [KM]

(1995) Evolution, the five-factor model, and levels of personality. *Journal of Personality* 63:525–67. [KM]

(1998) Evolution, culture, and the five-factor model. *Journal of Cross-Cultural Psychology* 29:119–49. [KM]

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

Macdonald, R. R. & Gilhooly, K. J. (1990) More about Linda or conjunctions in context. *European Journal of Cognitive Psychology* 2:57–70. [aKES]

Macintosh, N. J. (1998) *IQ and human intelligence.* Cambridge University Press. [EH]

Maher, P. (1993) *Betting on theories.* Cambridge University Press. [aKES]

Mandel, D. R. & Lehman, D. R. (1998) Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General* 127:269–85. [DRM]

Manktelow, K. I. & Evans, J. St. B. T. (1979) Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology* 70:477–88. [aKES]

Manktelow, K. I. & Over, D. E. (1991) Social roles and utilities in reasoning with deontic conditionals. *Cognition* 39:85–105. [aKES]

(1995) Deontic reasoning. In: *Perspectives on thinking and reasoning: Essays in honour of Peter Wason,* ed. S. E. Newstead & J. St. B. T. Evans. Erlbaum. [KIM]

Mansbridge, J. (1990) *Beyond self-interest.* University of Chicago Press. [VG]

March, J. G. (1988) Bounded rationality, ambiguity, and the engineering of choice. In: *Decision making: Descriptive, normative, and prescriptive interactions.* Cambridge University Press. [aKES]

Margolis, H. (1987) *Patterns, thinking, and cognition.* University of Chicago Press. [JEA, arKES]

Markovits, H. & Vachon, R. (1989) Reasoning with contrary-to-fact propositions. *Journal of Experimental Child Psychology* 47:398–412. [aKES]

Marr, D. (1982) *Vision.* W. H. Freeman. [arKES]

Massaro, D. W. (1994) A pattern recognition account of decision making. *Memory and Cognition* 22:616–27. [SS-F]

Matarazzo, J. D. (1972) *Wechsler's measurement and appraisal of adult intelligence* (fifth edition). Williams & Wilkins. [aKES]

McCabe, K., Rassenti, S. & Smith, V. (1996) Game theory and reciprocity in some extensive form games. *Proceedings of the National Academy of Sciences USA* 13428:13421. [RAM]

McCain, R. A. (1992) *A framework for cognitive economics.* Praeger. [RAM]

McCarthy, R. A. & Warrington, E. K. (1990) *Cognitive neuropsychology: A clinical introduction.* Academic Press. [rKES]

McClennan, E. F. (1990) *Rationality and dynamic choice.* Cambridge University Press. [PA]

McDowell, J. J. & Dallery, J. (1999) Falsification of matching theory: Changes in the symptote of Hernstein's hyperbola as a function of water deprivation. *Journal of the Experimental Analysis of Behavior* 72:251–68. [AJG]

McGeorge, P., Crawford, J. & Kelly, S. (1997) The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:239–45. [aKES]

McGraw, K. O. (1978) The detrimental effects of reward on performance: A literature review and a prediction model. In: *The hidden costs of reward,* ed. M. R. Lepper & D. Greene. Erlbaum. [AK]

McKelvey, R. D. & Palfrey, T. R. (1995) Quantal response equilibria in normal form games. *Games and Economic Behavior* 10:6–38. [DJZ]

(1998) Quantal response equilibria in extensive form games. *Experimental Economics* 1:9–41. [DJZ]

McKenzie, C. R. (1994) The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology* 26:209–39. [UH]

Messer, W. S. & Griggs, R. A. (1993) Another look at Linda. *Bulletin of the Psychonomic Society* 31:193–96. [aKES]

Miller, D. T., Turnbull, W. & McFarland, C. (1990) Counterfactual thinking and social perception: Thinking about what might have been. In: *Advances in experimental social psychology,* ed. M. P. Zanna. Academic Press. [aKES]

Miller, P. M. & Fagley, N. S. (1991) The effects of framing, problem variations, and providing rationale on choice. *Personality and Social Psychology Bulletin* 17:517–22. [aKES]

Morier, D. M. & Borgida, E. (1984) The conjunction fallacy: A task specific phenomenon? *Personality and Social Psychology Bulletin* 10:243–52. [aKES]

Morris, M., Sim, D. & Girotto, V. (1998) Distinguishing sources of cooperation in the one-round Prisoner's Dilemma. *Journal of Experimental Social Psychology* 34:494–512. [VG]

Morrison, F. J. (1997) Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology* 33:254–62. [rKES]

Morton, O. (1997) Doing what comes naturally: A new school of psychology finds reasons for your foolish heart. *The New Yorker* 73:102–107. [aKES]

Moshman, D. (1994) Reason, reasons, and reasoning: A constructivist account of human rationality. *Theory and Psychology* 4:245–60. [DM]

(1998) Cognitive development beyond childhood. In: *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th edition), D. Kuhn & R. Siegler. (Series editor, W. Damon). Wiley. [DM]

(1999) *Adolescent psychological development: Rationality, morality, and identity.* Erlbaum. [DM]

Moshman, D. & Franks, B. (1986) Development of the concept of inferential validity. *Child Development* 57:153–65. [arKES]

Moshman, D. & Geil, M. (1998) Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning* 4:231–48. [aKES]

Myerson, J., Rank, M. R., Raines, F. Q. & Schnitzler, M. A. (1998) Race and general cognitive ability: The myth of diminishing returns to education. *Psychological Science* 9:139–42. [rKES]

Nathanson, S. (1994) *The ideal of rationality.* Open Court. [arKES]

Navon, D. (1989a) The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system: I. Basic tenets. *European Journal of Cognitive Psychology* 1:191–213. [aKES]

Navon, D. (1989b) The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system: II. Application to the field of attention. *European Journal of Cognitive Psychology* 1:215–38. [aKES]

Neisser, U., Boodoo, G., Bouchard, T., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D., Loehlin, J., Perloff, R., Sternberg, R. & Urbina, S. (1996) Intelligence: Knowns and unknowns. *American Psychologist* 51:77–101. [arKES]

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

(1990) *Unified theories of cognition.* Harvard University Press. [aKES]

Newell, A. & Rosenbloom, P. S. (1981) Mechanisms of skill acquisition and the law of practice. In: *Cognitive skills and their acquisition,* ed. J. R. Anderson. Erlbaum. [SLS]

Newstead, S. E. & Evans, J. St. B. T., eds. (1995) *Perspectives on thinking and reasoning.* Erlbaum. [aKES]

Neyman, A. (1985) Bounded complexity justifies cooperation in the finitely repeated Prisoner's Dilemma. *Economics Letters* 19:227–29. [DJZ]

Nickerson, R. S. (1988) On improving thinking through instruction. In: *Review of research in education, vol. 15,* ed. E. Z. Rothkopf. American Educational Research Association. [rKES]

(1996) Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning* 2:1–31. [aKES]

(1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2:175–220. [aKES]

Niiniluoto, I. (1981) L. J. Cohen versus Bayesianism. *Behavioral and Brain Sciences* 4:349. [SO]

Nisbett, R. E. (1981) Lay arbitration of rules of inference. *Behavioral and Brain Sciences* 4:349–50. [aKES]

(1993) *Rules for reasoning.* Erlbaum. [rKES]

Nisbett, R. E. & Ross, L. (1980) *Human inference: Strategies and shortcomings of social judgment.* Prentice-Hall. [rKES]

Nisbett, R. E. & Wilson, T. (1977) Telling more than we know: Verbal reports on mental processes. *Psychological Review* 84:231–59. [rKES]

Norris, S. P. (1992) Testing for the disposition to think critically. *Informal Logic* 14:157–64. [rKES]

Nozick, R. (1993) *The nature of rationality.* Princeton University Press. [rKES]

Oaksford, M. & Chater, N. (1992) Bounded rationality in taking risks and drawing inferences. *Theory and Psychology* 2:225–30. [rKES]

(1993) Reasoning theories and bounded rationality. In: *Rationality: Psychological and philosophical perspectives,* ed. K. Manktelow & D. Over. Routledge. [MO, arKES]

(1994) A rational analysis of the selection task as optimal data selection. *Psychological Review* 101:608–31. [LJB, KIM, aKES]

(1995) Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning* 1:121–52. [MO, arKES]

(1996) Rational explanation of the selection task. *Psychological Review* 103:381–91. [aKES]

(1998) *Rationality in an uncertain world: Essays on the cognitive science of human reasoning.* Psychology Press. [MO, arKES]

Oaksford, M., Chater, N. & Grainger, B. (1999) Probabilistic effects in data selection. *Thinking and Reasoning* 5:193–243. [KO]

Oaksford, M., Chater, N., Grainger, B. & Larkin, J. (1997) Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:441–58. [aKES]

Oaksford, M., Chater, N. & Larkin, J. (in press) Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory and Cognition.* [MO]

Oatley, K. (1992) *Best laid schemes: The psychology of emotions.* Cambridge University Press. [arKES]

Oberauer, K., Wilhelm, O. & Rosas. D. R. (1999) Bayesian rationality for the selection task? A test of optimal data selection theory. *Thinking and Reasoning* 5:115–44. [KO]

Obiols, J. E., Serrano, F., Caparros, B., Subira, S. & Barrantes, N. (1999) Neurological soft signs in adolescents with poor performance on the continuous performance test: Markers of liability for schizophrenia spectrum disorders? *Psychiatry Research* 86:217–28. [MO]

O'Brien, D. P. (1995) Finding logic in human reasoning requires looking in the right places. In: *Perspectives on thinking and reasoning,* ed. S. E. Newstead & J. St. B. T. Evans. Erlbaum. [aKES]

Ohlsson, S. (1984) Induced strategy shifts in spatial reasoning. *Acta Psychologica* 57:47–67. [MB]

Olson, D. R. (1986) Intelligence and literacy: The relationship between intelligence and the technologies of representation and communication. In: *Practical intelligence,* ed. R. J. Sternberg & R. K. Wagner. Cambridge University Press. [aKES]

Osherson, D. N. (1995) Probability judgment. In: *Thinking, vol. 3,* ed. E. E. Smith & D. N. Osherson. MIT Press. [aKES]

Over, D. E. & Evans, J. St. B. T. (1997) Two cheers for deductive competence. *Current Psychology of Cognition* 16:255–78. [DEO]

Over, D. E. & Manktelow, K. I. (1993) Rationality, utility, and deontic reasoning. In: *Rationality: Psychological and philosophical perspectives,* ed. K. I. Manktelow & D. E. Over. Routledge. [KIM]

Overton, W. F. (1985) Scientific methodologies and the competence-moderator performance issue. In: *Moderators of competence,* ed. E. D. Neimark, R. DeLisi & J. L. Newman. Erlbaum. [aKES]

(1990) Competence and procedures: Constraints on the development of logical reasoning. In: *Reasoning, necessity, and logic,* ed. W. F. Overton. Erlbaum. [aKES]

Overton, W. F. & Newman, J. L. (1982) Cognitive development: A competence-activation/utilization approach. In: *Review of human development,* ed. T. Field, A. Huston, H. Quay, L. Troll & G. Finley. Wiley. [AK]

Pacini, R. & Epstein, S. (1999) The relation of rational and experiential information processing styles to personality, basic beliefs and the ration-bias phenomenon. *Journal of Personality and Social Psychology* 76:927–87. [SEN]

Parfit, D. (1984) *Reasons and persons.* Oxford University Press. [rKES]

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1992) Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology* 43:87–131. [AK]

Pearson, K. (1897) *The chances of death and other studies in evolution, vol. 1.* Arnold. [AJG]

Pelham, B. W., Sumarta, T. T. & Myaskovsky, L. (1994) The easy path from many to much: The numerosity heuristic. *Cognitive Psychology* 26:103–33. [LJB]

Pennington, B. F. & Ozonoff, S. (1996) Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry* 37:51–87. [rKES]

Perkins, D. N., Farady, M. & Bushey, B. (1991) Everyday reasoning and the roots of intelligence. In: *Informal reasoning and education,* ed. J. Voss, D. Perkins & J. Segal. Erlbaum. [aKES]

Perkins, D. N. & Grotzer, T. A. (1997) Teaching intelligence. *American Psychologist* 52:1125–33. [rKES]

Phillips, L. D. & Edwards, W. (1966) Conservatism in a simple probability inference task. *Journal of Experimental Psychology* 72:346–54. [aKES]

Phillips, L. D., Hays, W. L. & Edwards, W. (1966) Conservatism in complex probabilistic inference. *IEEE Transactions on Human Factors in Electronics* 7:7–18. [aKES]

Piaget, J. (1953) *Logic and psychology.* University of Manchester Press. [VFR]

Piattelli-Palmarini, M. (1994) *Inevitable illusions: How mistakes of reason rule our minds.* Wiley. [arKES]

Pinker, S. (1989) *Learnability and cognition: The acquisition of argument structure.* MIT Press. [SLS]

(1997) *How the mind works.* Norton. [WTD, aKES]

Plotkin, H. C. (1988) Behavior and evolution. In: *The role of behavior in evolution,* ed. H. C. Plotkin. MIT Press. [rKES]

Plous, S. (1993) *The psychology of judgment and decision making.* McGraw-Hill. [aKES]

Politzer, G. & Nguyen-Xuan, A. (1992) Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, relevance judgments or pragmatic schemas? *Quarterly Journal of Experimental Psychology* 44A:401–12. [KIM]

Politzer, G. & Noveck, I. A. (1991) Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research* 20:83–103. [aKES]

Pollock, J. L. (1991) OSCAR: A general theory of rationality. In: *Philosophy and AI: Essays at the interface,* ed. J. Cummins & J. L. Pollock. MIT Press. [arKES]

(1995) *Cognitive carpentry: A blueprint for how to build a person.* MIT Press. [arKES]

Popper, K. R. (1962) Why are the calculi of logic and arithmetic applicable to reality? In: *Conjectures and refutations: The growth of scientific knowledge,* ed. K. R. Popper. Basic Books. [JB]

Posner, M. I. & Snyder, C. R. R. (1975) Attention and cognitive control. In: *Information processing and cognition: The Loyola Symposium,* ed. R. L. Solso. Wiley. [arKES]

Potts, R. (1998) Variability selection in hominid evolution. *Evolutionary Anthropology* 7:81–96. [KM]

Quartz, S. R. & Sejnowski, T. J. (1998) The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences* 20:537–96. [rKES]

Quayle, J. D. & Ball, L. J. (in press) Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology:* Section A. [LJB]

Quiggin, J. (1982) A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3:323–43. [XTW]

Quine, W. (1960) *Word and object.* MIT Press. [rKES]

Rawls, J. (1971) *A theory of justice.* Harvard University Press. [JB]

Reber, A. S. (1992a) An evolutionary context for the cognitive unconscious. *Philosophical Psychology* 5:33–51. [rKES]

(1992b) The cognitive unconscious: An evolutionary perspective. *Consciousness and Cognition* 1:93–133. [rKES]

(1993) *Implicit learning and tacit knowledge.* Oxford University Press. [arKES]

Reber, A. S., Wakenfeld, F. F. & Hernstadt, R. (1991) Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:888–96. [aKES]

Reeves, T. & Lockhart, R. S. (1993) Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology: General* 122:207–26. [aKES]

Rescher, N. (1988) *Rationality: A philosophical inquiry into the nature and rationale of reason.* Oxford University Press. [aKES]

Resnik, M. D. (1987) *Choices: An introduction to decision theory.* University of Minnesota Press. [aKES]

Reyna, V. F. (1991) Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review* 11:317–36. [RH, VFR]

(1992) Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In: *Development of long-term retention,* ed. M. L. Howe, C. J. Brainerd & V. F. Reyna. Springer-Verlag. [VFR]

(1995) Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In: *New perspectives on interference and inhibition in cognition,* ed. F. N. Dempster & C. J. Brainerd. Academic Press. [PAK, VFR]

(1996) Conceptions of memory development, with implications for reasoning and decision making. *Annals of Child Development* 12:87–118. [VFR]

Reyna, V. F. & Brainerd, C. J. (1990) Fuzzy processing in transitivity development. *Annals of Operations Research* 23:37–63. [VFR]

(1991) Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making* 4:249–62. [VFR]

(1993) Fuzzy memory and mathematics in the classroom. In: *Memory in everyday life,* ed. G. M. Davies & R. H. Logie. North-Holland. [VFR]

(1994) The origins of probability judgment: A review of data and theories. In: *Subjective probability,* ed. G. Wright & P. Acton. Wiley. [VFR]

(1995) Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences* 7:1–75. [PAK, VFR]

(1998) Fuzzy-trace theory and false memory: New frontiers. *Journal of Experimental Child Psychology* 71:194–209. [VFR]

Reyna, V. F. & Ellis, S. C. (1994) Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science* 5:275–79. [PAK, VFR]

Reyna, V. F., Lloyd, F. J. & Brainerd, C. J. (in press) Memory, development, and rationality: An integrative theory of judgment and decision making. In: *Emerging perspectives on decision research,* ed. S. Schneider & J. Shanteau. Cambridge University Press. [VFR, aKES]

Rips, L. J. (1989) The psychology of knights and knaves. *Cognition* 31:85–116. [VG]

(1994) *The psychology of proof.* MIT Press. [aKES]

Rips, L. J. & Conrad, F. G. (1983) Individual differences in deduction. *Cognition and Brain Theory* 6:259–85. [aKES]

Roberts, M. J. (1993) Human reasoning: Deduction rules or mental models, or both? *Quarterly Journal of Experimental Psychology* 46A:569–89. [aKES]

Rogoff, B. & Lave, J., eds. (1984) *Everyday cognition: Its origin in social context.* Harvard University Press. [PAK]

Rose, S. (1998) *Lifelines: Biology beyond determinism.* Oxford University Press. [KS]

Rosenthal, R. & Rosnow, R. L. (1991) *Essentials of behavioral research: Methods and data analysis* (second edition). McGraw-Hill. [aKES]

Ross, L., Amabile, T. & Steinnetz, J. (1977) Social roles, social control, and biases in the social perception process. *Journal of Personality and Social Psychology* 36:485–94. [aKES]

Roth, A. & Erev, I. (1995) Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8:164–212. [DJZ]

Sá, W., West, R. F. & Stanovich, K. E. (1999) The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology* 91:497–510. [JF, arKES]

Sally, D. (1995) Conversation and cooperation in social dilemmas: Experimental evidence from 1958 to 1992. *Rationality and Society* 7:58–92. [VG]

Samuels, R., Stich, S. P. & Tremoulet, P. D. (1999) Rethinking rationality: From bleak implications to Darwinian modules. In: *What is cognitive science?,* ed. E. Lepore & Z. Pylyshyn. Blackwell. [rKES]

Savage, L. J. (1954) *The foundations of statistics.* Wiley. [arKES]

Schelling, T. (1960) *The strategy of conflict.* Harvard University Press. [RAM]

Schick, F. (1987) Rationality: The third dimension. *Economics and Philosophy* 3:49–66. [arKES]

(1997) *Making choices: A recasting of decision theory.* Cambridge University Press. [aKES]

Schober, M. F. & Clark, H. H. (1989) Understanding by addressees and overhearers. *Cognitive Psychology* 21:211–32. [SLS]

Scholnick, E. K. & Wing, C. S. (1995) Logic in conversation: Comparative studies of deduction in children and adults. *Cognitive Development* 10:319–45. [DM]

Schommer, M. (1990) Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology* 82:498–504. [rKES]

(1993) Epistemological development and academic performance among secondary students. *Journal of Educational Psychology* 85:406–11. [rKES]

(1994) Synthesizing epistemological belief research: Tentative understandings and provocative confusions. *Educational Psychology Review* 6:293–319. [rKES]

Schooler, C. (1984) Psychological effects of complex environments during the life span: A review and theory. *Intelligence* 8:259–81. [PAK]

Schwarz, N. (1996) *Cognition and communication: Judgmental biases, research methods, and the logic of conversation.* Erlbaum. [aKES]

Schweder, R. A. (1987) Comments on Plott and on Kahneman, Knetsch, and Thaler. In: *Rational choice: The contrast between economics and psychology,* ed. R. M. Hogarth & M. W. Reder. Chicago University Press. [aKES]

Scribner, S. & Cole, M. (1981) *The psychology of literacy.* Harvard University Press. [aKES]

Sedlmeier, P. (1997) Basic Bayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments, and Computers* 29:328–36. [aKES]

Seeley, T. D. (1994) Honey bee foragers as sensory units of their colonies. *Behavioral Ecology and Sociobiology* 34:51–62. [AJG]

Selten, R. (1975) Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4:22–55. [DJZ]

Sen, A. (1977) Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs* 6:317–44. [DJZ]

Shafir, E. (1993) Intuitions about rationality and cognition. In: *Rationality: Psychological and philosophical perspectives,* ed. K. Manktelow & D. Over. Routledge. [rKES]

(1994a) Uncertainty and the difficulty of thinking through disjunctions. *Cognition* 50:403–30. [aKES]

(1994b) Intransitivity of preferences in honey-bees: Support for comparative-evaluation of foraging options. *Animal Behaviour* 48:55–67. [PA]

Shafir, E. & Tversky, A. (1992) Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology* 24:449–74. [VG]

(1995) Decision making. In: *Thinking, vol. 3,* ed. E. E. Smith & D. N. Osherson. MIT Press. [aKES]

Shallice, T. (1988) *From neuropsychology to mental structure.* Cambridge University Press. [rKES]

Shanks, D. R. (1995) Is human learning rational? *Quarterly Journal of Experimental Psychology* 48A:257–79. [aKES]

Shanks, D. R. & St. John, M. F. (1994) Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* 17:367–447. [DJZ]

Shapiro, B. J. (1983) *Probability and certainty in seventeenth-century England.* Princeton University Press. [RH]

Shiffrin, R. M. & Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review* 84:127–90. [SLS, arKES]

Shultz, T. R., Fisher, G. W., Pratt, C. C. & Ruff, S. (1986) Selection of causal rules. *Child Development* 57:143–52. [MB]

Sieck, W. & Yates, J. F. (1997) Exposition effects on decision making: Choice and confidence in choice. *Organizational Behavior and Human Decision Processes* 70:207–19. [aKES]

Siegler, R. S. (1995) How does change occur? A microgenetic study of number conservation. *Cognitive Psychology* 28:225–73. [AJG]

Simon, H. A. (1956) Rational choice and the structure of the environment. *Psychological Review* 63:129–38. [PA, DH, aKES, XTW]

(1957) *Models of man.* Wiley. [aKES]

(1983) *Reason in human affairs.* Stanford University Press. [DH, aKES]

(1990) Invariants of human behavior. *Annual Review of Psychology* 41:1–19. [XTW]

(1992) *Economics, bounded rationality and the cognitive revolution.* Elgar. [PA]

Skinner, B. F. (1981) Selection by consequences. *Science* 213:501–504. [AJG]

Skyrms, B. (1986) *Choice and chance: An introduction to inductive logic* (third edition). Wadsworth. [aKES]

(1996) *The evolution of the social contract.* Cambridge University Press. [aKES]

Slatkin, M. (1974) Hedging one's evolutionary bets. *Nature* 250:705. [AJG]

Sloman, A. (1993) The mind as a control system. In: *Philosophy and cognitive science,* ed. C. Hookway & D. Peterson. Cambridge University Press. [rKES]

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

(1999) Rational versus arational models of thought. In: *The nature of cognition,* ed. R. J. Sternberg. MIT Press. [rKES]

Sloman, S. A. & Rips, L. J. (1998) Similarity as an explanatory construct. In: *Similarity and symbols in human thinking,* ed. S. A. Sloman & L. J. Rips. MIT Press. [rKES]

Slovic, P. (1995) The construction of preference. *American Psychologist* 50:364–71. [aKES]

Slovic, P., Fischhoff, B. & Lichtenstein, S. (1977) Behavioral decision theory. *Annual Review of Psychology* 28:1–39. [aKES]

Slovic, P. & Tversky, A. (1974) Who accepts Savage's axiom? *Behavioral Science* 19:368–73. [PA, DH, JJ, SLS, arKES]

Slugoski, B. R. & Wilson, A. E. (1998) Contribution of conversation skills to the production of judgmental errors. *European Journal of Social Psychology* 28:575–601. [aKES]

Smith, E. E., Patalino, A. L. & Jonides, J. (1998) Alternative strategies of categorization. In: *Similarity and symbols in human thinking,* ed. S. A. Sloman & L. J. Rips. MIT Press. [rKES]

Smith, S. M. & Levin, I. P. (1996) Need for cognition and choice framing effects. *Journal of Behavioral Decision Making* 9:283–90. [arKES]

Smith, V. L. & Walker, J. M. (1993) Monetary rewards and decision cost in experimental economics. *Economic Inquiry* 31:245–61. [AK]

Snyderman, M. & Rothman, S. (1990) *The IQ controversy: The media and public policy.* Transaction. [aKES]

Sober, E. & Wilson, D. S. (1994) A critical review of philosophical work on the units of selection problem. *Philosophy of Science* 61:534–55. [rKES]

Sohal, V. S. & Hasselmo, M. E. (1998) Changes in GABA-$_{(B)}$ modulation during a theta cycle may be analogous to the fall of temperature during annealing. *Neural Computation* 10:869–82. [AJG]

Sophian, C. (1997) Beyond competence: The significance of performance for conceptual development. *Cognitive Development* 12:281–303. [rKES]

Spearman, C. (1904) General intelligence, objectively determined and measured. *American Journal of Psychology* 15:201–93. [aKES]

(1927) *The abilities of man.* Macmillan. [aKES, RJS]

Stahl, D. O. (1996) Boundedly rational rule learning in a guessing game. *Games and Economic Behavior* 16:303–30. [DJZ]

(1998) Is step-j thinking an arbitrary modelling restriction or a fact of human nature? *Journal of Economic Behavior and Organization* 27:33–51. [DJZ]

Stankov, L. & Dunn, S. (1993) Physical substrata of mental energy: Brain capacity and efficiency of cerebral metabolism. *Learning and Individual Differences* 5:241–57. [aKES]

Stanovich, K. E. (1990) Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review* 10:72–100. [rKES]

(1993) Does reading make you smarter? Literacy and the development of verbal intelligence. In: *Advances in child development and behavior, vol. 24,* ed. H. Reese. Academic Press. [rKES]

(1999) *Who is rational? Studies of individual differences in reasoning.* Erlbaum. [JF, arKES]

(2000) *Progress in understanding reading: a Scientific foundations and new frontiers.* Guilford Press. [rKES]

Stanovich, K. E., Nathan, R., West, R. F. & Vala-Rossi, M. (1985) Children's word recognition in context: Spreading activation, expectancy, and modularity. *Child Development* 56:1418–28. [rKES]

Stanovich, K. E. & West, R. F. (1979) Mechanisms of sentence context effects: Automatic activation and conscious attention. *Memory and Cognition* 7:77–85. [rKES]

(1983) On priming by a sentence context. *Journal of Experimental Psychology: General* 112:1–36. [rKES]

(1997) Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology* 89:342–57. [JF, SEN, arKES]

(1998a) Cognitive ability and variation in selection task performance. *Thinking and Reasoning* 4:193–230. [arKES]

(1998b) Individual differences in framing and conjunction effects. *Thinking and Reasoning* 4:289–317. [RH, JJ, aKES]

(1998c) Individual differences in rational thought. *Journal of Experimental Psychology: General* 127:161–88. [VG, JK, arKES]

(1998d) Who uses base rates and P(D/-H)? An analysis of individual differences. *Memory and Cognition* 28:161–79. [aKES]

(1999) Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology* 38:349–85. [DH, JJ, arKES]

Stein, E. (1996) *Without good reason: The rationality debate in philosophy and cognitive science.* Oxford University Press. [arKES]

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

(1990) *Metaphors of mind: Conceptions of the nature of intelligence.* Cambridge University Press. [rKES]

(1997a) *Successful intelligence.* Plume. [RJS]

(1997b) The concept of intelligence and its role in lifelong learning and success. *American Psychologist* 52:1030–37. [aKES]

(1997c) *Thinking styles.* Cambridge University Press. [rKES]

Sternberg, R. J. & Detterman, D. K. (1986) *What is intelligence?* Ablex. [rKES]

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J., Wagner, R. K., Williams, W. M., Snook, S. & Grigorenko, E. L. (2000) *Practical intelligence.* Cambridge University Press. [RJS]

Sternberg, R. J. & Gardner, M. K. (1982) A componential interpretation of the general factor in human intelligence. In: *A model for intelligence,* H. J. Eysenck. Springer-Verlag. [aKES]

Sternberg, R. J. & Kaufman, J. C. (1998) Human abilities. *Annual Review of Psychology* 49:479–502. [aKES]

Sternberg, R. J. & Wagner, R. K. (1993) The g-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science* 2:1–5. [ASG]

Stich, S. P. (1990) *The fragmentation of reason*. MIT Press. [arKES]

Stich, S. P. & Nisbett, R. E. (1980) Justification and the psychology of human reasoning. *Philosophy of Science* 47:188–202. [arKES]

Stolarz-Fantino, S., Fantino, E. & Kulik, J. (1996) The conjunction fallacy: Differential incidence as a function of descriptive frames and educational context. *Contemporary Educational Psychology* 21:208–18. [SS-F, DJZ]

Takemura, K. (1992) Effect of decision time on framing of decision: A case of risky choice behavior. *Psychologia* 35:180–85. [aKES]

(1993) The effect of decision frame and decision justification on risky choice. *Japanese Psychological Research* 35:36–40. [aKES]

(1994) Influence of elaboration on the framing of decision. *Journal of Psychology* 128:33–39. [aKES]

Tan, H.-T. & Yates, J. F. (1995) Sunk cost effects: The influence of instruction and future return rates. *Organizational Behavior and Human Decision Processes* 63(3):311–19. [DH]

Teigen, K. H. (1988) When are low-probability events judged to be "probable"? Effects of outcome-set characteristics on verbal probability judgments. *Acta Psychologica* 67:27–38. [KHT]

(1999) When equal chances = good chances: Verbal probabilities and the equiprobability effect. (submitted). [KHT]

Teigen, K. H. & Brun, W. (1999) The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes* 80:155–90. [KHT]

Thagard, P. (1982) From the descriptive to the normative in philosophy and logic. *Philosophy of Science* 49:24–42. [aKES]

(1992) *Conceptual revolutions.* Princeton University Press. [aKES]

Thaler, R. H. (1992) *The winner's curse: Paradoxes and anomalies of economic life.* Free Press. [arKES]

Thompson, V. A. (1994) Interpretational factors in conditional reasoning. *Memory and Cognition* 22:742–58. [MO]

Thompson, V. A. & Mann, J. M. (1995) Perceived necessity explains the dissociation between logic and meaning: The case of "Only if." *Journal of Experimental Psychology: Learning, Memory and Cognition* 21:1554–67. [MO]

Todd, M. & Gigerenzer, G. (2000) Simple heuristics that make us smart. *Behavioral and Brain Sciences* 23(5). [DEO]

Tooby, J. & Cosmides, L. (1992) The psychological foundations of culture. In: *The adapted mind: Evolutionary psychology and the generation of culture,* ed. J. H. Barkow, L. Cosmides & J. Tooby. Oxford University Press. [KM, DEO]

Tschirgi, J. E. (1980) Sensible reasoning: A hypothesis about hypothesis. *Child Development* 51:1–10. [aKES]

Tsotsos, J. K. (1991) Computational resources do constrain behavior: *Behavioral and Brain Sciences* 14:506–507. [JJ]

Tversky, A. (1969) Intransitivity of preferences. *Psychological Review* 76:31–48. [PA]

(1975) A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis* 9:163–73. [arKES]

Tversky, A. & Kahneman, D. (1977) Causal thinking in judgment under uncertainty. In: *Basic problems in methodology and linguistics,* ed. R. Butts & J. Hintikka. Reidel. [SO]

(1981) The framing of decisions and the psychology of choice. *Science* 211:453–58. [aKES]

(1982) Evidential impact of base rates. In: *Judgment under uncertainty: Heuristics and biases,* ed. D. Kahneman, P. Slovic & A. Tversky. Cambridge University Press. [aKES]

(1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90:293–315. [DK, VFR, aKES, SS-F]

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

(1993) *Biological approaches to the study of human intelligence.* Ablex. [aKES]

Verplanken, B. (1993) Need for cognition and external information search: Responses to time pressure during decision-making. *Journal of Research in Personality* 27:238–52. [aKES]

Wagenaar, W. A. (1972) Generation of random sequences by human subjects: A critical survey of the literature. *Psychological Bulletin* 77:65–72. [aKES]

Wang, X. T. (1996) Domain-specific rationality in human choices: Violations of utility axioms and social contexts. *Cognition* 60:31–63. [XTW]

Wang, X. T., Simons, F. & Bredart, S. (in press) Social cues and verbal framing in risky choice. *Journal of Behavioral Decision Making.* [XTW]

Warnick, J. & Slonin, R. L. (1999) Learning to trust: Unobserved multi-period behavioral strategies using observed stage game actions. Paper presented at the Public Choice Society and Economic Science Association joint meeting, New Orleans, March 12- 14, 1999. [DJZ]

Wason, P. C. (1966) Reasoning. In: *New horizons in psychology,* ed. B. Foss. Penguin. [aKES]

Wason, P. C. & Evans, J. St. B. T. (1975) Dual processes in reasoning? *Cognition* 3:141–54. [DEO, rKES]

Wasserman, E. A., Dorner, W. W. & Kao, S. F. (1990) Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Motivation, and Cognition* 16:509–21. [aKES]

Weber, E. U. (1994) From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin* 115:228–42. [XTW]

Wenger, M. J. & McKinzie, D. I. (1996) Ontogenetic differences in variability on simple measures of learning: Theoretical and practical implications. *Developmental Psychobiology* 29:219–39. [AJG]

Werner, H. (1948) *Comparative psychology of mental development.* Humanities Press. [VFR]

West, R. F. & Stanovich, K. E. (1978) Automatic contextual facilitation in readers of three ages. *Child Development* 49:717–27. [rKES]

(1986) Robust effects of syntactic structure on visual word processing. *Memory and Cognition* 14:104–12. [rKES]

Wetherick, N. E. (1971) Representativeness in a reasoning problem: A reply to Shapiro. *Bulletin of the British Psychological Society* 24:213–14. [arKES]

(1993) Human rationality. In: *Rationality: Psychological and philosophical perspectives,* ed. K. Manktelow & D. Over. Routledge. [aKES]

(1995) Reasoning and rationality: A critique of some experimental paradigms. *Theory and Psychology* 5:429–48. [aKES]

Wigdor, A. K. & Garner, W. R., eds. (1982) *Ability testing: Uses, consequences, and controversies.* National Academy Press. [RJS]

Willingham, D. B. (1998) A neuropsychological theory of motor-skill learning. *Psychological Review* 105:558–84. [rKES]

(1999) The neural basis of motor-skill learning. *Current Directions in Psychological Science* 8:178–82. [rKES]

Wilson, T. D. & Brekke, N. (1994) Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin* 116:117–42. [JF, rKES]

Wilson, T. D., Dunn, D. S., Kraft, D. & Lisle, D. J. (1989) Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. *Advances in Experimental Social Psychology* 22:287–342. [JF]

Wilson, T. D. & Schooler, J. W. (1991) Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology* 60:181–92. [DH, rKES]

Windschitl, P. D. & Wells, G. L. (1996) Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied* 2:343–64. [KHT]

(1998) The alternative-outcomes effect. *Journal of Personality and Social Psychology* 75:1411–23. [KHT]

Winer, G. A., Craig, R. K. & Weinbaum, E. (1992) Adults failure on misleading weight-conservation tests – A developmental analysis. *Developmental Psychology* 28:109–20. [VFR]

Wolfe, C. (1995) Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory account. *Journal of Behavioral Decision Making* 8:85–108. [VFR]

Yaari, M. E. (1987) The dual theory of choice under risk. *Econometrica* 55:95–115. [XTW]

Yalch, R. F. & Elmore-Yalch, R. (1984) The effect of numbers on the route to persuasion. *Journal of Consumer Research* 11:522–27. [JF]

Yates, J. F. & Carlson, B. W. (1986) Conjunction errors: Evidence for multiple judgment procedures, including “signed summation.” *Organizational Behavior and Human Decision Processes* 37:230–53. [SS-F]

Yates, J. F, Lee, J. & Shinotsuka, H. (1996) Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes* 65:138–47. [aKES]

Yoshimura, J. & Clark, C. W. (1991) Individual adaptations in stochastic environments. *Evolutionary Ecology* 5:173–92. [AJG]

Zizzo, D. J. (1999) How do people learn to be rational? Unpublished manuscript. [DJZ]

Zizzo, D. J., Stolarz-Fantino, S., Wen, J. & Fantino, E. (2000) A violation of the monotonicity axiom: Experimental evidence on the conjunction fallacy. *Journal of Economic Behavior and Organization* 41:263–76. [SS-F, DJZ]