Cognitive Psychology **38,** 349–385 (1999)
Article ID cogp.1998.0700, available online at http://www.idealibrary.com on IDEAL®

# Discrepancies Between Normative and Descriptive Models of Decision Making and the Understanding/Acceptance Principle

Keith E. Stanovich

*University of Toronto*

and

Richard F. West

*James Madison University*

Several tasks from the heuristics and biases literature were examined in light of Slovic and Tversky's (1974) understanding/acceptance principle—that the deeper the understanding of a normative principle, the greater the tendency to respond in accord with it. The principle was instantiated both correlationally and experimentally. An individual differences version was used to examine whether individuals higher in tendencies toward reflective thought and in cognitive ability would be more likely to behave normatively. In a second application of the understanding/acceptance principle, subjects were presented with arguments both for and against normative choices and it was observed whether, on a readministration of the task, performance was more likely to move in a normative direction. Several discrepancies between performance and normative models could be explained by the understanding/acceptance principle. However, several gaps between descriptive and normative models (particularly those deriving from some noncausal base rate problems) were not clarified by the understanding/acceptance principle—they could not be explained in terms of varying task understanding or tendencies toward reflective thought. The results demonstrate how the variation and instability in responses can be

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Keith E. Stanovich and a James Madison University Program Faculty Assistance Grant to Richard F. West. Preparation of the manuscript was supported by a Connaught Research Fellowship in the Social Sciences to Keith E. Stanovich. Jonathan Baron, Michael Doherty, and an anonymous reviewer are thanked for their comments on an earlier version of the manuscript. The authors thank Walter Sa and Maggie Toplak for assistance in data coding.

Address correspondence and reprint requests to Keith E. Stanovich, Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education, University of Toronto, 252 Bloor Street West, Toronto, Ontario, Canada M5S 1V6. E-mail: KStanovich@oise.utoronto.ca.



0010-0285/99 $30.00

analyzed to yield inferences about why descriptive and normative models of human reasoning and decision making sometimes do not coincide. 

The literature on reasoning and decision making contains many demonstrations that human performance often deviates from the behavior considered normative under certain models of optimal response. For example, people violate the axioms of utility theory, they do not always obey the probability calculus, and they often fail to employ basic principles of logic (for summaries of the large literature, see Baron, 1994b; Dawes, 1988; Evans, 1989; Evans & Over, 1996; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1979; Newstead & Evans, 1995; Piattelli-Palmarini, 1994; Raiffa, 1985; Shafir & Tversky, 1995). However, the theoretical interpretation of these discrepancies between descriptive models and normative models in the heuristics and biases literature continues to be a source of considerable contention (e.g., Baron, 1994a; Cohen, 1981, 1983; Evans & Over, 1996; Gigerenzer, 1996; Kahneman & Tversky, 1983, 1996; Stanovich, 1999; Stein, 1996). Some investigators view such normative–descriptive discrepancies as demonstrations of systematic human irrationality. Others (e.g., Cohen, 1981) deny that such discrepancies are indicative of human irrationality. Instead, these critics argue that the task has been misconceived by the experimenter and that an inappropriate normative model has been applied (the inappropriate norm argument, see Gigerenzer, 1991, 1996; Stanovich, 1999). In cases where the appropriate normative model is not in doubt, critics have argued that participants have a different construal of the problem and are responding normatively to a different problem (the differential construal argument, see Adler, 1984, 1991; Hilton, 1995; Oaksford & Chater, 1994; Stanovich, 1999).

When speaking of gaps between normative and descriptive models, it is in fact more accurate to say that there is a gap between *mean* or *modal* performance and the normative response, because there is considerable response variability displayed on most tasks. Even on tasks where the modal response is nonnormative, some subjects do give the normative response. It is argued here that the nature of these individual differences has implications for explanations of the gap between normative and descriptive models. Specifically, critiques of the irrationality assumption in the early heuristics and biases literature in terms of inappropriate norms or differential construals are undermined if the finding that spawned the critiques—that typical human performance departs from a specified normative model—displays systematic lability and/or is predictable from individual difference variables that reflect cognitive competence and depth of problem understanding.

## REFLECTIVE EQUILIBRIUM AND THE UNDERSTANDING/ ACCEPTANCE PRINCIPLE

Consider Cohen's (1981) attempt to show that human irrationality cannot be experimentally demonstrated. He notes that in linguistics, normative principles derive from intuitions about grammaticality. These intuitions about grammaticality are in turn based on our linguistic competence. Thus, in linguistics, normative principles are directly indexed to linguistic competence and the two cannot come apart. Cohen (1981) imports this argument into the domain of human reasoning by proposing, analogously, that normative principles about what constitutes good reasoning come from our intuitions about what constitutes good reasoning and that our intuitions about what constitutes good reasoning come from our reasoning competence. Thus, as in linguistics, normative principles of reasoning are directly indexed to reasoning competence and the two cannot come apart. The normative models of rational thought simply *are* our thought, and humans are—by definition—rational. No experimentation is necessary to establish it (see Stein, 1996 and Stanovich, 1999, for fuller discussions and critiques).

In order to justify the two premises of his argument—that (1) normative principles about what constitutes good reasoning come from our intuitions about what constitutes good reasoning and (2) our intuitions about what constitutes good reasoning come from our reasoning competence—Cohen (1981) employs the concept of reflective equilibrium (Daniels, 1979; Elgin, 1996; Goodman, 1965; Rawls, 1971). Reflective equilibrium is achieved by a process of constraint satisfaction whereby: ''rules and particular inferences alike are justified by being brought into agreement with each other. A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences'' (Goodman, 1965, p. 64). Thus, Cohen's view is transformed into the notion that ''the normative principles of reasoning come from a process of reflective equilibrium with our intuitions about what constitutes good reasoning as input'' (Stein, 1996, p. 142) and that our intuitions about what constitutes good reasoning come from our reasoning competence.

These are the premises that guarantee perfect human rationality according to Cohen (1981). Thus, any replicable gaps between descriptive and normative models must be due to some error on the experimenter's part—the subject not being characterized by any systematic irrationalities. As noted previously, it is typically argued that either the experimenter has applied the wrong normative model to the task or, if it is allowed that the experimenter may be applying the correct normative model to the problem, it is posited that the subject has construed the problem in some other way and is providing a normatively appropriate answer to a different problem.

Cohen (1981) and other critics (e.g., Gigerenzer, 1996; Macdonald, 1986; Wetherick, 1995) are at pains to show that any normative/descriptive gaps can be so explained. But what are the gaps that are purported to be explained by these critiques? They consist of modal or mean performance departing from the response deemed normative under the model being applied. They are gaps that are uninformed by any analysis of the degree of understanding of the subject. Specifically, it is assumed that these nonnormative responses are in reflective equilibrium. In the present series of studies we exam this assumption.

The framework for the present research derives from an argument first advanced by Slovic and Tversky (1974). In their 1974 article, they speculated on a ''mock'' debate between Allais and Savage about the independence axiom of utility theory (Allais, 1953; Savage, 1954). 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 observed that ''the deeper the understanding of the axiom, the greater the readiness to accept it'' (pp. 372–373). Slovic and Tversky (1974) argued that this understanding/acceptance congruence suggests that the gap between the descriptive and normative was due to an initial failure to fully process and/or understand the task.

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 at least in part to factors that prevented problem understanding (for example, confusing problem presentation by the experimenter or lack of ability or reflectiveness on the part of the subject). Such a finding would increase our confidence in the normative appropriateness of the axioms for the particular problem under consideration. In contrast, if better understanding failed to result in greater acceptance of the axiom, then the argument that its application to a particular situation might be inappropriate would be strengthened. Using the understanding/acceptance principle in this manner represents a pretheoretical commitment to a type of naturalistic epistemology that is widely endorsed in decision science (Larrick, Nisbett, & Morgan, 1993; March, 1988; Slovic, 1995; Stanovich, 1999; Thagard & Nisbett, 1983).

Using their understanding/acceptance principle, Slovic and Tversky (1974) found little support for the independence axiom. When presented with arguments to explicate both the Allais and Savage 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 and MacCrimmon & Larsson, 1979 for more mixed results on the independence axiom using related paradigms). In contrast, Doherty, Schiavo, Tweney, and Mynatt (1981) found that manipulations which increased the understanding of diagnosticity in the Bayesian sense markedly reduced the tendency toward endorsing pseudodiagnostic behavior (ignoring $P(D/{\sim}H)$ in information search).

Although Slovic and Tversky (1974) failed to find support for the particular normative application that they examined, they did present a principle that may be of more general usefulness in theoretical debates about why human performance deviates from normative models. The operational tool revealed by Slovic and Tversky's (1974) development of the understanding/acceptance principle is that the direction that performance moves in response to increased understanding provides an empirical clue as to what is the proper normative interpretation for the particular problem.

In the present paper we generalized their understanding/acceptance principle to a variety of tasks drawn from the heuristics and biases literature that have spawned debate about the reasons for the discrepancies between normative and descriptive models. We employed two different methods for examining 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. As previously mentioned, they found little support for the normative principle of independence—although using related methods MacCrimmon (1968; MacCrimmon & Larsson, 1979) did find more (see also Doherty et al., 1981). Other methods of manipulating understanding have provided consistent evidence in favor of applications 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 reduces framing effects (Larrick, Smith, & Yates, 1992; Miller & Fagley, 1991; Sieck & Yates, 1997; Takemura, 1992, 1993, 1994).

As an alternative to manipulating understanding, the understanding/acceptance principle can be transformed into an individual differences prediction. For example, we might interpret the principle as indicating that more reflective and engaged reasoners are more likely to respond in accord with normative principles. Thus, we might expect that those individuals with cognitive/personality characteristics more conducive to deeper understanding will be more accepting of the normative principles of reasoning and decision. Larrick et al. (1993) presented just such an argument in their analysis of cost-benefit reasoning—claiming that superior cognitive ability should be associated with the endorsement of the axioms of instrumental rationality: ''Intelligent people [should] 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). They found empirical support for this prediction.

Another cognitive/personality variable that is a marker for reflective and engaged reasoning is need for cognition—a dispositional construct encompassing the tendency toward thoughtful analysis, intellectual engagement, the desire for understanding, and the tendency toward thorough information processing (Ackerman & Heggestad, 1997; Cacioppo, Petty, Feinstein, & Jarvis, 1996). Consistent with the individual differences prediction of the understanding/acceptance principle, Smith and Levin (1996) found that framing effects (indicating violations of descriptive invariance) were smaller among individuals higher in need for cognition. Interestingly, even critics of the heuristics and biases research tradition have endorsed the individual differences version of the understanding/acceptance principle as a means adjudicating arguments about the proper application of normative models (Cohen, 1982; Funder, 1987; Lopes & Oden, 1991; Wetherick, 1971, 1995; see discussion in Stanovich, 1999).

In the present series of studies, we examined several tasks from the heuristics and biases literature in which performance often deviates from that considered optimal according to standard normative models. We employed three methods to assess whether the understanding/acceptance principle appeared to justify the normative model traditionally applied to the task. The first method was argument explication similar to that employed in Experiment 1 of Slovic and Tversky (1974). Subsequent to answering the problems for the first time, subjects were presented with arguments explicating normative and nonnormative principles favoring a given response. Subjects evaluated these arguments. At a later point in the experiment, the subject responded again to the same problem. We then looked for asymmetries in the percentages of subjects who persevered with their response pattern and those who changed their responses subsequent to the explications—specifically whether there was a greater proportion of subjects shifting in the normative direction.

In addition to the argument presentation method, we conducted two individual differences analyses on each task—one on a measure of cognitive ability (as in Larrick et al., 1993) and one on the thinking disposition of need for cognition (Cacioppo et al., 1996). We will thus examine whether, for each task, reasoners with high cognitive ability and high tendencies toward thorough information processing are more likely to endorse applications of normative principles that have been controversial in the heuristics and biases literature.

## PARTICIPANTS AND GENERAL METHOD

All of the problems examined here were presented to 663 participants. Subjects were undergraduate students (206 males and 457 females) recruited through an introductory psychology subject pool at a medium-sized state university. Their mean age was 18.5 years ($SD$ = 1.0). For many of the problems described below, a few subjects failed to respond, and thus the reported sample sizes are lower than 663. Similarly, some participants were missing cognitive ability scores and thus analyses involving this variable always contain less than 663.

The measure of cognitive ability used in this investigation was the subject's SAT (Scholastic Aptitude Test) total score, a measure that loads highly on psychometric g (Carroll, 1993; Matarazzo, 1972). Students were asked to indicate their verbal and mathematical SAT scores on a demographics sheet. The mean reported total SAT score was 1185 (*SD* = 104). This mean reported SAT total score was reasonably close to the institutional means of 1192 for 1994–1995 and 1179 for 1995–1996. The SAT scores of these subjects will be reported as rescaled scores since the majority were derived from testings subsequent to the April 1995 rescaling of the SAT. Scores from testings prior to the April 1995 rescaling have been recentered according to ETS formulas.

The 18-item need for cognition scale (Cacioppo et al., 1996) was administered to all subjects. The scale has been examined in over 200 studies (see Cacioppo et al., 1996) and has consistently been associated with greater intellectual engagement, more thorough information processing and information search, and higher levels of reflection. Examples of need for cognition items are: ''I really enjoy a task that involves coming up with new solutions to problems,'' ''I like to have the responsibility of handling a situation that requires a lot of thinking,'' and ''I like tasks that require little thought once I've learned them,'' the latter reverse scored. All the items are listed in the Appendix of Cacioppo et al. (1996). Subjects responded on a six-point scale: 1 = Disagree Strongly, 2 = Disagree Moderately, 3 = Disagree Slightly, 4 = Agree Slightly, 5 = Agree Moderately, 6 = Agree Strongly. Nine of the items are reverse scored. The reliability of the scale in published research is high, and in this sample the split-half reliability (Spearman–Brown corrected) was .89. The 18 items of the need for cognition scale were intermixed with a large number of items from other scales that served as fillers. The total score for the 18 items was used in the analyses that follow. The mean score on the scale was 68.0 (*SD* = 12.3). One subject did not complete the need for cognition scale.

Participants completed the problems during a single two-hour session in which they also completed some other tasks not part of the present investigation. They were tested in small groups of 3-4 individuals. The problems were interspersed between other unrelated tasks. Thus, in conditions where problems were presented twice, there was approximately one hour between administrations of the problem.

## EXAMPLES OF APPLICATIONS OF NORMATIVE PRINCIPLES REAFFIRMED BY AN APPLICATION OF THE UNDERSTANDING/ACCEPTANCE ASSUMPTION

We will first illustrate several cases where application of the understanding/acceptance principle reaffirmed the normative interpretation of the task that is traditionally applied. In the next section, we explore cases where applications of the understanding/acceptance principle call into question traditional task interpretations.

*The Selection Task: Attention to the Denominator of the Likelihood Ratio*

Two deviations from normatively correct Bayesian reasoning have been the focus of much research. The two deviations are most easily characterized if Bayes' rule is expressed in the ratio form where the odds favoring the focal hypothesis (H) are derived by multiplying the likelihood ratio of the observed datum (D) by the prior odds favoring the focal hypothesis:

$$\frac{P(H/D)}{P(\sim H/D)} = \frac{P(D/H)}{P(D/\sim H)} \times \frac{P(H)}{P(\sim H)}.$$

The first deviation is the tendency to ignore—or at least to pay insufficient attention to—the denominator of the likelihood ratio, P(D/∼H), the probability of the datum given that the focal hypothesis is false (Beyth-Marom & Fischhoff, 1983; Doherty, Chadwick, Garavan, Barr, & Mynatt, 1996; Doherty, Mynatt, Tweney, & Schiavo, 1979; Einhorn & Hogarth, 1978; Wasserman, Dorner,& Kao, 1990; Wolfe, 1995). The other deviation from Bayesian reasoning that has been the subject of intense investigation is the tendency for individuals to underweight the prior probability, P(H), when it is presented as a statistical base rate—and especially when it has no apparent causal relevance to the focal hypothesis (Bar-Hillel, 1980, 1984; 1990; Fischhoff & Bar-Hillel, 1984; Koehler, 1996; Lyon & Slovic, 1976; Tversky & Kahneman, 1982).

We used a variant of the selection paradigm studied by Doherty and Mynatt (1990) to examine both deviations from Bayesian reasoning and we will focus in this section on the first deviation—the tendency to underestimate the relevance of P(D/∼H). Participants 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 estimate the probability that the patient has the disease ''Digirosa''? Below are 4 pieces of information that may or may not be relevant to determining the probability. Please indicate *all* of the pieces of information that are necessary to determine the probability, but *only* those pieces of information that are necessary to do so.'' Participants 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. Because P(H) and P(∼H) are complements, only three pieces of information are necessary to calculate the posterior probability. However, P(D/∼H) clearly must be selected because it is a critical component of the likelihood ratio.

Four patterns accounted for over 93% of the choices. The normatively correct choice of P(H), P(D/H), and P(D/∼H) was made by 9.4% of the sample. The most popular choice in the sample was the two components of

TABLE 1
Mean SAT and Need for Cognition Scores of Subjects Making the Normative and Nonnormative Choices on Various Tasks (Number of Subjects in Parentheses)

| Variable | Nonnormative | Normative | *t* value |
|---|---|---|---|
| Selection Task, P(D/∼H) | | | |
| SAT Total | 1173 (296) | 1196 (350) | 2.89*** |
| Need for Cognition | 67.9 (304) | 68.2 (356) | 0.34 |
| Selection Task, P(H) | | | |
| SAT Total | 1187 (432) | 1183 (214) | −0.39 |
| Need for Cognition | 67.2 (444) | 69.7 (216) | 2.46** |
| Sunk Cost Problem | | | |
| SAT Total | 1173 (262) | 1194 (385) | 2.54** |
| Need for Cognition | 66.4 (267) | 69.1 (394) | 2.80*** |
| Newcomb's Problem | | | |
| SAT Total | 1181 (405) | 1192 (242) | 1.26 |
| Need for Cognition | 67.1 (415) | 69.5 (246) | 2.40** |
| Prisoner's Dilemma | | | |
| SAT Total | 1176 (264) | 1191 (383) | 1.83 |
| Need for Cognition | 68.3 (272) | 67.8 (389) | −0.59 |

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

the likelihood ratio, P(D/H) and P(D/∼H)—the response of 38.6% of the sample. One-fifth of the sample (19.5%) 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). Finally, 25.7% of the sample chose P(D/H) only.

On an individual component basis, almost all subjects (94.9%) viewed P(D/H) as relevant and very few (2.1%) viewed P(∼H) as relevant. Overall, 54.0% of our participants chose P(D/∼H) and 32.8% of the sample thought it was necessary to know the base rate, P(H). More than 26% of the sample viewed *neither* the base rate nor P(D/∼H) as relevant.

Table 1 presents the mean SAT scores of the subjects who chose P(D/∼H) as one of their selections and those who did not. The scores of the former were significantly higher. However, the next row of Table 1 indicates that this group was not significantly higher in need for cognition.

Subsequent to completing the selection task, 109 of our participants were asked to evaluate the following argument in favor of choosing P(D/∼H): ''It is important that we know the % of people with*out* Digirosa who have a red rash. This is important because if the % of people both *with* and *without* Digirosa who have the red rash were the same, then we would know that the presence of the red rash would not help distinguish between those with and without Digirosa.'' Subjects rated this argument on a six-point scale ranging from ''extremely weak'' to ''extremely strong.'' Subsequent to com-

TABLE 2

Percentage of Subjects Who Changed Their Responses on the Second Administration of the Problem after Evaluating the Arguments for and Against Choosing P(D/∼H) in the Selection Task

| | | One normative argument | |
|---|---|---|---|
| Initial response | N | % Changing responses | |
| Nonnormative | 45 | 48.9 | $\chi^2(1) = 19.49$*** |
| Normative | 64 | 10.9 | phi = .42 |
| | | One nonnormative argument | |
| Initial response | N | % Changing responses | |
| Nonnormative | 50 | 22.0 | $\chi^2(1) = 0.05$ |
| Normative | 59 | 23.7 | phi = −.02 |
| | | Both arguments | |
| Initial response | N | % Changing responses | |
| Nonnormative | 206 | 34.0 | $\chi^2(1) = 10.79$*** |
| Normative | 221 | 19.9 | phi = .16 |

* $p < .05$.
** $p < .025$.
*** $p < .01$.

pleting several further problems (some reported below) and scales, participants were presented again with the selection task and other tasks (discussed below) which they had completed before. These problems were preceded by the following instructions: ‘‘You have seen some of these problems before, but we would like you to consider them again. Please *disregard* your previous response to it. Examine the problem again with a fresh eye, perhaps considering the arguments that you evaluated. Make what you *currently* think is the best response.’’ The subjects then completed for the second time an identical version of the Doherty and Mynatt (1990) selection task.

Our analyses here focus on the choice of the P(D/∼H) alternative and the issue of whether subjects changed their responses from the first administration of the task to the second after seeing an argument explicating a reason for the normative response. On this task, response change could be in one of two directions: making the normative response [choosing P(D/∼H)] on the first administration and making the nonnormative response [failing to choose P(D/∼H)] on the second administration; or making the nonnormative choice on the first administration and the normative choice on the second. As the first analysis in Table 2 indicates, nearly half (48.9%) of the subjects giving the nonnormative response on the first administration chose P(D/∼H) when presented again with the task subsequent to evaluating an argument favoring the normative choice. Because only 10.9% of the subjects originally giving the normative response changed in the other direction, there was a significant tendency for responses to change in the normative direction.

With our next group of 109 subjects we attempted to determine whether the movement would be as strong in the nonnormative direction if subjects evaluated a single argument *against* choosing P(D/∼H). The argument was the following: ‘‘The % of people with*out* Digirosa who have a red rash is irrelevant. The probability to be determined is the probability that the person *has* Digirosa and the key piece of evidence that is needed for that probability is the % of people with Digirosa who have a red rash. If this % is high, then it is necessarily true that it is probable that this person has Digirosa.’’ In the same sequence as the previous group, this group of subjects evaluated this argument immediately after the first administration of the selection task and then subsequent to completing several further problems were presented again with the selection task preceded by the same instructions reproduced above. The second analysis in Table 2 indicates that 23.7% of the subjects in this condition moved in the nonnormative direction, whereas 22.0% moved in the normative direction—a difference that was not statistically significant.

A third group of 427 subjects was given *both* arguments to evaluate[1] after the first administration of the selection task (214 receiving the normative argument first and 213 receiving the normative argument second). As the third analysis in Table 2 indicates, after being presented with both arguments, significantly more subjects moved in the normative direction than in the nonnormative direction (34.0% versus 19.9%). Thus, when presented with arguments explicating reasons for both choices, subjects were more responsive to the argument for the normative choice—a result consistent with the greater potency displayed by the normative argument when only one argument was presented. These results based on the manipulation of argument explication are consistent with the finding that subjects higher in cognitive ability also were more likely to choose P(D/∼H).

### *The Selection Task: Choosing the Base Rate*

A parallel set of analyses were carried out on the base rate choice, P(H), in the selection task. The second set of analyses presented in Table 1 indicates that there was not a significant difference in SAT scores between those who chose P(H) and those who did not. However, those choosing the base rate were significantly higher in need for cognition.

Subsequent to completing the selection task and evaluating a single argument relevant to P(D/∼H), 109 subjects received the following argument in favor of choosing the base rate: ‘‘The % of people with Digirosa is needed to determine the probability because, if Digirosa is very infrequent in the population and some people *without* Digirosa *also* have red rashes, then the probability of Digirosa might still be low even if the person has a red rash.’’ Subjects rated this argument on the same six-point scale.

[1] In a later discussion (see Table 10) we deal with the issue of the relative strength of the two arguments.

TABLE 3

Percentage of Subjects Who Changed Their Responses on the Second Administration of the Problem after Evaluating the Arguments for and Against Choosing P(H) in the Selection Task

| | | One normative argument | |
|---|---|---|---|
| Initial response | N | % Changing responses | |
| Nonnormative | 69 | 39.1 | $\chi^2(1) = 6.99$*** |
| Normative | 40 | 15.0 | phi = .25 |
| | | One nonnormative argument | |
| Initial response | N | % Changing responses | |
| Nonnormative | 83 | 15.7 | $\chi^2(1) = 2.91$ |
| Normative | 26 | 30.8 | phi = −.16 |
| | | Both arguments | |
| Initial response | N | % Changing responses | |
| Nonnormative | 283 | 27.2 | $\chi^2(1) = 14.52$*** |
| Normative | 144 | 11.1 | phi = .18 |

* $p < .05$.
** $p < .025$.
*** $p < .01$.

The first analysis in Table 3 focuses on whether subjects changed their response on the second administration of the task after evaluating a single argument explicating reasons for this choice. More than 39% of the individuals who originally responded nonnormatively changed their responses in the normative direction, compared with 15.0% who changed in the other direction, a difference that was statistically significant. The next analysis in Table 3 presents the parallel outcome when 109 different subjects were given the following argument for the irrelevance of the base rate: ''The percentage of people with Digirosa is irrelevant because this particular patient has a red rash, and thus the % of people who have Digirosa is not needed when trying to determine the probability that someone has Digirosa given that they have a red rash.'' Here, although there was a larger shift in the nonnormative than in the normative direction, the difference was not statistically significant.

A third group of 427 subjects was given *both* arguments to evaluate after the first administration of the selection task (213 receiving the normative argument first and 214 receiving the normative argument second). As the third analysis in Table 3 indicates, after being presented with both arguments, significantly more subjects moved in the normative direction than in the nonnormative direction (27.2% versus 11.1%). Thus, when presented with arguments explicating reasons for both choices, subjects were more responsive to the argument for the normative choice. These results based on the manipulation of argument explication are consistent with the understanding/acceptance principle—as is the finding that subjects higher in need for cognition were also more likely to choose P(H).

*Honoring Sunk Costs*

The next problem was designed to reveal susceptibility to the tendency to honor sunk costs. A person is said to be mistakenly honoring sunk costs when they persist in a nonoptimal activity because previous resources have been spent on the activity (for numerous examples, see Arkes, 1996; Arkes & Blumer, 1985; Bornstein & Chapman, 1995; Frisch, 1993; Thaler, 1980). The tendency to honor sunk costs is viewed as irrational economic behavior under traditional analyses, because actions should be determined only by future consequences, not past expenditures (Baron, 1993, 1994a, 1998). Our sunk cost problem was drawn from Frisch (1993). Subjects read the following two scenarios:

> Situation X: You are staying alone in a hotel room. You have paid $6.95 to see a movie on pay TV. After 15 minutes you are bored, the movie seems pretty bad, and there are other things on regular TV to watch that might be more enjoyable. Would you continue to watch the movie or not?
> a. I would continue to watch the movie
> b. I would switch to another channel
>
> Situation Y: You are staying alone in a hotel room. You turn on the TV and there is a movie on. After 15 minutes you are bored, the movie seems pretty bad, and there are other things on regular TV to watch that might be more enjoyable. Would you continue to watch the movie or not?
> a. I would continue to watch the movie
> b. I would switch to another channel

The participants responded to both situations which were presented adjacently and in the order given above for each subject. Overall, there was a large effect in the direction of honoring sunk costs; 42.7% of the sample thought they would watch the movie if they had paid for it, whereas only 2.4% of the sample thought they would watch the movie if they had not paid for it. An analysis of response patterns among individual subjects indicated that 268 honored sunk costs (they would watch the movie if they had paid for it but not if they had not paid for it), 394 subjects responded consistently (15 watching the movie in both cases and 379 not watching it in both cases), and one subject displayed a reverse effect (this subject was eliminated from the analyses that follow).

Table 1 indicates that the subjects displaying a sunk cost effect had significantly lower SAT scores than those responding normatively—a result that is convergent with the results of Larrick et al. (1993). Additionally, the subjects displaying a sunk cost effect were significantly lower in need for cognition than those responding normatively.

Subsequent to completing the sunk cost problem, 105 of the participants were asked to evaluate the following argument against honoring sunk costs: ‘‘The two situations are very similar. Once I have paid the $6.95, it is water under the bridge. Fifteen minutes after turning on the TV, I am in the same situation in both A and B: I am faced with the choice of whether it is more

TABLE 4
Percentage of Subjects Who Changed Their Responses on the Second Administration of the Sunk Cost Problem

| | | | |
|---|---|---|---|
| | | One normative argument | |
| Initial response | N | % Changing responses | |
| Nonnormative | 36 | 44.4 | $\chi^2(1) = 28.75$*** |
| Normative | 69 | 2.9 | phi = .52 |
| | | One nonnormative argument | |
| Initial response | N | % Changing responses | |
| Nonnormative | 45 | 15.6 | $\chi^2(1) = 1.54$ |
| Normative | 63 | 7.9 | phi = .12 |
| | | Both arguments | |
| Initial response | N | % Changing responses | |
| Nonnormative | 176 | 27.8 | $\chi^2(1) = 46.76$*** |
| Normative | 240 | 4.2 | phi = .34 |

* $p < .05$.
** $p < .025$.
*** $p < .01$.

enjoyable to continue watching the movie or to watch regular TV. I can't get the $6.95 back, so I should forget about the money. If my enjoyment is increased by turning off the movie and watching regular TV, I should turn it off whether or not I paid for the movie. Continuing to watch the movie just because I paid $6.95 is like throwing good time after bad money. The only thing that matters is doing the thing that is most enjoyable *now*, regardless of how I arrived at the present situation.'' Subjects rated this argument on the same six-point scale.

The first analysis in Table 4 focuses on whether subjects changed their response pattern on the second administration of the task after evaluating a single argument explicating reasons for avoiding sunk costs. More than 44% of those who initially responded nonnormatively changed their responses in the normative direction, compared with 2.9% who changed in the other direction, a difference that was statistically significant. The next analysis in Table 4 indicates that presenting an argument in favor of honoring sunk costs was ineffective in causing changes in the nonnormative direction. The argument utilized one of the most common rationales for honoring sunk costs—that failing to do so entails waste (see Arkes, 1996)—and was as follows: ''The two situations are very different. If I had paid $6.95 for the movie and then didn't watch it, I would have wasted $6.95. If I did watch the movie, even if I was bored and didn't like it, I would have gotten something for my money so it wouldn't be a total waste. Whereas, if I had paid nothing for the movie and turned it off, nothing would be wasted.'' Even after seeing this argument, more subjects moved in the normative direction than the reverse (15.6% versus 7.9%).

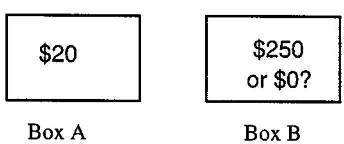
A third group of 416 subjects was given *both* arguments to evaluate after the first administration of the sunk cost task (209 receiving the normative argument first and 207 receiving the normative argument second). As the third analysis in Table 4 indicates, after being presented with both arguments, significantly more subjects moved in the normative direction than in the non-normative direction (27.8% versus 4.2%). Thus, when presented with arguments explicating reasons for both choices, subjects were much more responsive to the argument for the normative choice—avoiding honoring sunk costs. This result is consistent with the greater potency displayed by the normative argument when only one argument was presented and with the fact that the subjects responding normatively were higher in cognitive ability and need for cognition.

To get an idea of the maximum range of the effect of variability in understanding, we conducted a comparison of the two outlying groups in terms of argument understanding. In order to illustrate the impact of all three of these factors (argument evaluation, cognitive ability, need for cognition) on normative choices we dichotomized the sample based on their SAT scores and need for cognition scores. Analyses subsequent to partitioning the sample based on median splits indicated the following. Among those 172 subjects in the lower half of the distributions of both cognitive ability and need for cognition, 82 honored sunk costs (47.7%) on the first administration of the task. In contrast, after evaluating both arguments, only 31 of the 102 subjects (30.4%) in the upper half of the distributions of both cognitive ability and need for cognition honored sunk costs. Thus, the combined effects of evaluating arguments both for and against the normative choice, high cognitive ability, and high need for cognition resulted in a 36.3% decrease (17.3/47.7) in the proportion of the sample violating the normative stricture on this problem.

We next turn our attention to two problems where there has been intense philosophical dispute regarding what should be considered the normative response.

### *Newcomb's Problem*

Newcomb's problem, introduced into the literature by Nozick (1969), has been the subject of intense philosophical dispute (Gibbard & Harper, 1978; Hurley, 1991; Nozick, 1993; see Campbell & Sowden, 1985 for a collection of useful readings). The version given to 662 subjects in the present study was modeled on Shafir and Tversky (1992) and was as follows:



> Here is a problem that asks you to make use of your imagination. Consider the two boxes above. Imagine Box A contains $20 for sure. Imagine that Box B may contain either $250 or nothing. Pretend your options will be to:
>
> 1. Choose both Boxes A and B (and collect the money that is in both boxes).
> 2. Choose Box B only (and collect only the money that is in Box B).
>
> Imagine now that we have a computer program called the ''Predictor'' that has analyzed the pattern of the responses you have already made to all of the earlier questions. Based on this analysis, the program has already predicted your preference for this problem and has already loaded the boxes accordingly. If, based on this analysis of your previous preferences, the program has predicted that you will take both boxes, then it has left Box B empty. On the other hand, if it has predicted that you will take only Box B, then it has already put $250 in that box. So far, the program has been very successful: Most of the participants who chose Box B received $250; in contrast, few of those who chose both boxes found $250 in Box B.
> Which of the above options would you choose?
>
> a) Choose both Box A and Box B.
> b) Choose Box B only.

The choices in Newcomb's problem pit the dictates of evidential decision theory against those of dominance within the framework of causal decision theory (Campbell, 1985; Eells, 1982; Gibbard & Harper, 1978; Hurley, 1991; Nozick, 1993; Resnik, 1987). The two-box consequentialist choice is clearly viewed as normative by most psychologists who have examined the problem (Shafir, 1994; Shafir & Tversky, 1992), especially in this version which was specifically designed by Shafir and Tversky (1992) to emphasize that the predictor's choice had already been made and to remove some seemingly supernatural elements from the original formulation of the problem (Nozick, 1969). Yet despite these changes in the problem which make the dominant consequentialist choice quite salient, 65% of Shafir and Tversky's (1992) subjects preferred to take only one box—although their sample was somewhat small (only 40 subjects).

Using a sample size which was an order of magnitude larger, we replicated their result—63% of our sample preferred the one box option on the first administration of the problem. Table 1 indicates that the subjects choosing the two-box alternative had higher SAT scores than those choosing the one-box alternative, but this difference was not significant. However, those choosing two boxes did have significantly higher need for cognition scores.

Subsequent to completing the first administration of the problem, 109 of the participants were asked to evaluate the following consequentialist argument (see Nozick, 1969) in favor of taking both boxes: ''The Predictor has already made the prediction and has already either put the $250 in Box B or has not. If the Predictor has already put the $250 in Box B, and I take both boxes, I get $250 + $20, whereas if I take only Box B, I get only $250. If the Predictor has not put the $250 in Box B, and I take both boxes, I get $20, whereas if I take only Box B I get no money. Therefore, whether the $250 is there or not, I get $20 more by taking both boxes rather than only

TABLE 5
Percentage of Subjects Who Changed Their Responses on the Second Administration of Newcomb's Problem

| Initial response | N | % Changing responses | |
|---|---|---|---|
| | | One normative argument | |
| Nonnormative | 67 | 35.8 | $\chi^2(1) = 11.39$*** |
| Normative | 42 | 7.1 | phi = .32 |
| | | One nonnormative argument | |
| Initial response | N | % Changing responses | |
| Nonnormative | 71 | 5.6 | $\chi^2(1) = 13.24$*** |
| Normative | 42 | 31.0 | phi = −.34 |
| | | Both arguments | |
| Initial response | N | % Changing responses | |
| Nonnormative | 276 | 24.3 | $\chi^2(1) = 19.21$*** |
| Normative | 160 | 7.5 | phi = .21 |

* $p < .05$.
** $p < .025$.
*** $p < .01$.

Box B. So I should take both boxes.'' Subjects rated this argument on the same six-point scale.

The first analysis in Table 5 indicates that 35.8% of the participants changed their responses in the normative direction, compared with 7.1% who changed in the other direction, a difference that was statistically significant. However, the next analysis indicates that the movement was just as strong in the other direction when another group of subjects had to evaluate a single argument in favor of taking only one box. This argument, adapted from Nozick (1969), was the following: ''If I choose both boxes, the Predictor almost certainly would have predicted this and would not have put the $250 in Box B, and so therefore I will get only $20. If I take only Box B, the Predictor almost certainly would have predicted this and will have put the $250 in Box B, and so therefore I will get the $250. Thus, if I take both boxes, I almost certainly will get only $20. But, if I take just Box B, I almost certainly will get $250. Therefore, I should choose just Box B.''

A third group of 436 subjects was given *both* arguments to evaluate after the first administration of the problem (217 receiving the normative argument first and 219 receiving the normative argument second). As the third analysis in Table 5 indicates, after being presented with both arguments, significantly more subjects moved in the normative (two-box) direction than in the nonnormative direction (24.3% versus 7.5%). Thus, when presented with arguments explicating reasons for both choices, subjects were more responsive to the argument for the normative choice.

Again, to explore the combined impact of all three factors (argument evaluation, cognitive ability, need for cognition) a post hoc analysis was conducted and revealed the following. Among those 130 subjects in the lower half of the distributions of both cognitive ability and need for cognition, only 41 chose two boxes (31.5%) on the first administration of the task. In contrast, after evaluating both arguments, 53.8% of the 106 subjects in the upper half of the distributions of both cognitive ability and need for cognition chose two boxes. Thus, the combined effects of evaluating arguments both for and against the normative choice, high cognitive ability, and high need for cognition resulted in a 70.8% increase (22.3/31.5) in the proportion of subjects making the normative two-box choice. However, the following (similarly contentious) problem indicates that this outcome does not always obtain.

## PROBLEMATIC APPLICATIONS OF THE UNDERSTANDING/ ACCEPTANCE PRINCIPLE

In all of the tasks examined so far, the standard normative model usually applied to the task has been reinforced by an application of the understanding/acceptance principle. We now present a series of analyses in which an application of the understanding/acceptance principle did not have this outcome.

### *The Prisoner's Dilemma*

The prisoner's dilemma has generated an enormous literature (e.g., Hargreaves Heap & Varoufakis, 1995; Rapoport & Chammah, 1965; Rasmusen, 1989; Skyrms, 1996). The version given to 662 subjects in the present study was modelled on Shafir and Tversky (1992) and was as follows:

Intercollegiate Computer Game

This game was originally designed to be played by pairs of students who were sitting in front of different computers on the same computer system. Since we are not using computers today, *please use your imagination*, and pretend that you are sitting in front of a computer and playing the Intercollegiate Computer Game with another student.

In this game you will be presented with a situation involving you and one other player who is sitting at a computer in another room. You cannot communicate with each other. The situation requires that you make a strategic decision: to cooperate or to compete with the other player. The other player will have to make the same decision.

The situation is represented by a payoff matrix that will determine how much money each of you earns depending on whether you compete or cooperate. The matrix looks like the following:

| | Other Cooperates | Other Competes |
|---|---|---|
| You Cooperate | You: 20<br>Other: 20 | You: 5<br>Other: 25 |
| You Compete | You: 25<br>Other: 5 | You: 10<br>Other: 10 |

> According to this matrix, if you both cooperate you will both earn $20 each. If you cooperate and the other person competes, the other will earn $25 and you will earn only $5. Similarly, if you compete and the other person cooperates, you will earn $25 and the other person will earn only $5. Finally, if you both choose to compete, you will each earn $10. Not knowing what the other person will choose to do, what would you choose?
> (a) I would choose to compete
> (b) I would choose to cooperate

Competing is the dominant action for each player in this situation because whatever the other player does, each person is better off competing than cooperating. The fact that this individually rational (dominant) action leads to an outcome that is suboptimal for both ($10, when each could have had $20 by cooperating) is what has piqued the interest of social scientists and philosophers in this problem. As Nozick (1993) puts it: ‘‘The combination of (what appears to be) their individual rationalities leads them to forgo an attainable better situation and thus is Pareto-suboptimal’’ (p. 50).

However, in the one-shot game (with no communication between players or repeated play) the compete strategy is usually championed as rational—sometimes using parallels between the logic of the prisoner’s dilemma and the logic underlying Newcomb’s problem. This connection between the two problems has been discussed by several authors (Campbell & Sowden, 1985; Hurley, 1991; Lewis, 1979; Nozick, 1993). For example, Shafir and Tversky (1992) point out that ‘‘in both cases, the outcome depends on the choice that you make and on that made by another being—the other player in the Prisoner’s Dilemma and the Predictor in Newcomb’s problem. In both cases one option (competing or taking both boxes) dominates the other, yet the other option (cooperating or taking just one box) seems preferable if the being—the Predictor or the other player—knows what you will do, or will act like you’’ (p. 460)—but in neither case can anything you do affect what the Predictor or the other player has already done. Nevertheless, Shafir and Tversky (1992) found a cooperation rate of 37% using a one-shot problem similar to that presented above.

In our much larger sample of participants, we observed a similar cooperation rate of 41%. Table 1 indicates that the subjects choosing to compete had higher SAT scores than cooperators but this difference did not reach significance. However, the difference in need for cognition was in the other direction—subjects choosing to compete had *lower* scores—although again the difference was not significant.

Subsequent to completing the first administration of the problem, 109 of the participants were asked to evaluate the following consequentialist argument in favor of competing: ‘‘No matter what the other person does, I am better off competing. If the other person cooperates and I compete, I get $25 rather than $20. If the other person competes and I compete, I get $10 rather than $5. Competing is always the better strategy for me, so it is the better choice.’’ Subjects rated this argument on the same six-point scale.

TABLE 6
Percentage of Subjects Who Changed Their Responses on the Second Administration of the Prisoner's Dilemma

| Initial response | N | % Changing responses | |
|---|---|---|---|
| **One normative argument** | | | |
| Nonnormative | 37 | 51.4 | $\chi^2(1) = 28.07$*** |
| Normative | 72 | 6.9 | phi = .51 |
| **One nonnormative argument** | | | |
| Nonnormative | 43 | 7.0 | $\chi^2(1) = 8.07$*** |
| Normative | 68 | 29.4 | phi = −.27 |
| **Both arguments** | | | |
| Nonnormative | 191 | 23.0 | $\chi^2(1) = 2.72$ |
| Normative | 245 | 16.7 | phi = .08 |

* $p < .05$.
** $p < .025$.
*** $p < .01$.

The first analysis in Table 6 indicates that when the task was readministered, over half of the cooperators (51.4%) changed their responses compared with only 6.9% of the competers who changed in the other direction, a difference that was statistically significant. However, the next analysis indicates that there was a strong (but not quite as large—phi coefficient of −.27 compared to .51) change in the other direction when another group of subjects had to evaluate a single argument in favor of cooperating. This argument is adapted from Hurley's (1991) discussion of the notion of collective action as the source of the attractiveness of the cooperative response: ''The rational thing for both of us to do is to both cooperate and get $20 rather than to both compete and get only $10. The other player probably realizes this, too, just like I do. Therefore, I should cooperate so that we both end up with $20, rather than $10.''

A third group of 436 subjects was given *both* arguments to evaluate after the first administration of the problem (218 receiving the normative argument first and 218 receiving the normative argument second). As the third analysis in Table 6 indicates, there was no significant difference in the proportions of subjects who moved in the normative direction compared with those who moved in the nonnormative direction. That the three factors (argument evaluation, cognitive ability, need for cognition) had minimal association with the tendency to give the compete response is indicated by the following analysis in which the individual differences variables were dichotomized. Among those 172 subjects in the lower half of the distributions of both cognitive ability and need for cognition, 102 (59.3%) chose to compete on the first administration of the task. This proportion is actually higher than the proportion of subjects in the upper half of the distributions of both cognitive ability and need for cognition who chose to compete after evaluating the two arguments (54.2%).

Deeper understanding of the logic of the prisoner's dilemma situation—as evidenced by considering arguments for both responses or by the tendency to engage in reflective thought or the ability to successfully engage in such thought—was not associated with the compete response that is traditionally considered normative in the one-shot problem. Of course, the traditional analysis assumes a purely selfish motivation (interestingly, economic training has been found to associate with noncooperative responses, see Frank, Gilovich, & Regan, 1993). Allowing other motivations to enter changes the normative model (Baron, 1994b). The failure of the understanding/acceptance principle to point strongly in one direction or the other on this problem might be indicating that there may be more than one normatively appropriate interpretation of this problem operative within this group of participants.

*Noncausal Base Rates: The Cab Problem*

So-called noncausal base rates—those bearing no obvious causal relationship to the criterion behavior—have been the subject of over a decade's worth of contentious dispute (Bar-Hillel, 1990; Cohen, 1979, 1981; Cosmides & Tooby, 1996; Gigerenzer, 1991, 1993, 1996; Gigerenzer & Hoffrage, 1995; Kahneman & Tversky, 1996; Koehler, 1996). We examined two such problems that are notorious for provoking contention. The first was the well-known cab problem (see Bar-Hillel, 1980; Lyon & Slovic, 1976; Tversky & Kahneman, 1982) that was presented to 644 subjects: ''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% of the cabs in the city are Green and 15% are Blue. A witness reported that the cab in the accident was 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 called about 80% of the Blue cabs blue, but called 20% of the Blue cabs green. The witness also called about 80% of the Green cabs green, but called 20% of the Green cabs blue. What is the probability (expressed as a percentage ranging from 0 to 100%) that the cab involved in the accident was Blue?''

It should be noted that this version of the cab problem clarified the witness identification phrase in a manner not done in the original version of this problem. The original wording, which was ''the witness correctly identified each of the two colors 80 percent of the time,'' might have encouraged subjects to confuse P(D/H) with P(H/D) and thus seem to be ignoring the base rate (see Braine, Connell, Freitag, & O'Brien, 1990; Dawes, Mirels, Gold, & Donahue, 1993; Macchi, 1995; Koehler, 1996). Thus, the version used in this study clarified the witness identification phrase.

For purposes of analysis, responses to this question were scored categori-

TABLE 7a
Mean Scores of Groups Classified as Indicant, Base Rate, and Bayesian on the Cab Problem

| | Indicant ($n$ = 252) | Base rate ($n$ = 214) | Bayesian ($n$ = 178) | $F$ ratio |
|---|---|---|---|---|
| SAT total | 1205[a] | 1174[b] | 1170[b] | 7.68*** |
| Need for cognition | 69.2 | 66.5 | 67.9 | 2.90* |

*Note.* $df$ = 2,628 for SAT, and 2,640 for Need for Cognition. Means with different superscripts ([a,b]) are significantly different.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

cally in terms of whether subjects relied on the indicant information, relied on the base rate, or amalgamated the base rate and indicant in a manner approximating Bayes rule (which yields .41 as the posterior probability of the cab being blue). Operationally, posterior probabilities greater than or equal to 70% were scored as reliance on the indicant information, probabilities less than or equal to 20% were scored as reliance on the base-rate information, and probabilities between 20% and 70% were interpreted as indicating Bayesian amalgamation. Using this classification scheme on the data from the first administration of the problem, 252 subjects were classified as reliant on the indicant (responses of ≥ 70%), 214 subjects were classified as reliant on the base rate (responses of ≤ 20%), and 178 were classified as approximately normatively Bayesian (responses between 20% and 70%).

Table 7a displays the mean scores of these three groups on the two cognitive/personality variables. There were significant differences in cognitive ability (as indicated by SAT scores) with the subjects in the indicant group having significantly higher SAT scores than either the base rate or the Bayesian groups who did not differ from each other. The indicant group

TABLE 7b
Mean Scores of Groups Classified as Indicant, Intermediate, and Bayesian on the Disease Problem

| | Indicant ($n$ = 318) | Intermediate ($n$ = 85) | Bayesian ($n$ = 242) | $F$ ratio |
|---|---|---|---|---|
| SAT total | 1182 | 1157[a] | 1202[b] | 6.38*** |
| Need for cognition | 67.8 | 68.1 | 68.4 | 0.21 |

*Note.* $df$ = 2,628 for SAT, and 2,642 for Need for Cognition. Means with different superscripts ([a,b]) are significantly different (Scheffe).
* $p < .10$.
** $p < .05$.
*** $p < .01$.

TABLE 8
Percentage of Subjects Who Changed Their Responses on the Second Administration of the Cab Problem

| Initial response | N | % Changing | % Changing to normative | % Changing to Nonnormative |
|---|---|---|---|---|
| One normative argument | | | | |
| Indicant | 46 | 43.5 | 26.1 | 17.4 |
| Bayesian | 23 | 47.8 | | 47.8 |
| Base rate | 30 | 30.0 | 23.3 | 6.7 |
| One nonnormative argument | | | | |
| Indicant | 38 | 10.5 | 2.6 | 7.9 |
| Bayesian | 33 | 36.4 | | 36.4 |
| Base rate | 32 | 46.9 | 18.8 | 28.1 |
| Both arguments | | | | |
| Indicant | 160 | 43.7 | 25.6 | 18.1 |
| Bayesian | 111 | 42.3 | | 42.3 |
| Base rate | 139 | 32.3 | 20.1 | 12.2 |

also attained the highest mean score on the need for cognition scale but the overall effect of group did not quite attain significance for this variable ($p =$ .056).

Subsequent to completing the cab problem, 99 of the participants were asked to evaluate the following argument designed to make the subjects who were highly reliant on indicant information pay more attention to the base rate: ‘‘The probability that the cab involved in the accident was Blue must be less than 50% because 20% of the time the witness calls Green cabs blue and Green cabs are very common (85% of all cabs). Thus, there will be more misidentified Green cabs than there will be correctly identified Blue cabs because only 15% of all cabs are Blue and only 80% of these will be identified correctly.’’ Subjects rated this argument on the same six-point scale. The first analysis in Table 8 indicates that when the task was readministered, 43.5% of the indicant subjects (20 of 46) did change their responses. However, almost half of these (8 of 20) overshot the mark and joined the base rate group (gave responses $\leq$ 20%). Interestingly, just as high a proportion (47.8%) of Bayesian subjects (11 of 23) changed their responses, 8 of the 11 moving in the direction of the base rate—the direction of the argument.

With our second group of 103 subjects we presented an argument designed to drive responses in the direction of the *indicant*: ‘‘The probability that the cab involved in the accident was Blue must be much greater than 50% because the witness’s accuracy in identifying cab colors (80%) is much greater than 50%.'' This argument caused 36.4% of the Bayesian subjects (12 of 33) to change their responses, 8 of 12 in the direction of the indicant.

Most important were the results from a third group of 410 who were given *both* arguments to evaluate after the first administration of the problem (206 receiving the base rate argument first and 204 receiving the indicant argument first). As the third analysis in Table 8 indicates, very similar proportions of indicant and Bayesian subjects (43.7% and 42.3%, respectively) changed their responses to one of the other categories. More specifically, 20.7% of the Bayesians moved to the indicant category and 25.6% of the indicant subjects moved to the Bayesian category.

Thus, the results regarding the cab problem were somewhat different from those concerning the selection of the base rate in the selection task. In the latter, high need for cognition was associated with selecting the base rate, and the base rate argument was more potent than the counter-base-rate argument. In the cab problem, cognitive ability and need for cognition were positively associated with reliance on the indicant, and the indicant argument was just as potent as the base rate argument. The vastly different task requirements might clearly be contributing to these differing patterns. In the selection task, the subject must simply choose the base rate when it is presented along with other useful information (the components of the likelihood ratio). In the cab problem, the base rate must actually be incorporated in a way that alters the final response which is a numerical probability assessment.

### *Noncausal Base Rates: The Disease Problem*

The second noncausal base rate problem was a disease testing problem modeled on Casscells, Schoenberger, and Graboys (1978): ''Imagine that disease X occurs in one in every 1000 people. A test has been developed to detect the disease. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, 5% of all people who are perfectly healthy test positive for the disease. Imagine that we have given this test to a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people. What is the chance (expressed as a percentage ranging from 0% to 100%) that a person found to have a positive result actually has the disease?''

The Bayesian posterior probability for this problem is slightly less than .02. Thus, responses of less than or equal to 10% were interpreted as indicating Bayesian amalgamation,[2] responses greater than or equal to 80% were

[2] A small proportion (12.4%) of the responses in this category were less than 1% and thus may have reflected total reliance on the base rate. As a result, this category includes both Bayesian and base rate responders. The results were unchanged when the base rate responders were removed.

scored as indicating strong reliance on indicant information, and responses between 10% and 80% were scored as intermediate. As in previous studies similar to this one (e.g., Cosmides & Tooby, 1996), the majority of subjects gave responses that were either Bayesian or that reflected strong reliance on indicant information: 318 were classified as strongly indicant, 242 as Bayesian, and 85 as intermediate. Table 7b presents the mean scores of these groups on the cognitive/personality variables. The Bayesian subjects had the highest mean SAT score—significantly higher than the mean SAT score of the intermediate subjects. However, the SAT scores of the indicant and Bayesian subjects did not differ significantly. The three groups did not differ significantly in need for cognition.

The analysis of responses to the argument evaluation component of the study will focus on the subjects classified as indicant and Bayesian (these two groups comprise over 85% of the subjects and the two arguments were framed to support these two extreme positions). Subsequent to completing the disease problem, 90 participants classified as either indicant or Bayesian were asked to evaluate the following argument designed to make subjects pay more attention to the base rate: ''The chance that a person with a positive test result has the disease must be low because the test *wrongly* indicates that 5% of the *healthy* people have the disease. Because there are many more healthy people than people with the disease, most of the people with positive test results will *not* have the disease.''

The first analysis in Table 9 indicates that after readministration of the task 43.8% of the indicant group changed their responses in the normative direction (were classified as either Bayesian or intermediate), compared with

TABLE 9
Percentage of Subjects Who Changed Their Responses on the Second Administration of the Disease Problem

| | | | |
|---|---|---|---|
| | | One normative argument | |
| Initial response | N | % Changing responses | |
| Nonnormative | 48 | 43.8 | $\chi^2(1) = 7.67$*** |
| Normative | 42 | 16.7 | phi = .29 |
| | | One nonnormative argument | |
| Initial response | N | % Changing responses | |
| Nonnormative | 50 | 7.0 | $\chi^2(1) = 11.24$*** |
| Normative | 44 | 36.4 | phi = −.35 |
| | | Both arguments | |
| Initial response | N | % Changing responses | |
| Nonnormative | 216 | 18.5 | $\chi^2(1) = 12.55$*** |
| Normative | 149 | 34.9 | phi = .19 |

* $p < .05$.
** $p < .025$.
*** $p < .01$.

just 16.7% of the Bayesian subjects who shifted in the other direction (were classified as either indicant or intermediate), a difference that was statistically significant. However, the next analysis indicates that the movement was just as strong in the other direction (phi coefficient of −.35 versus .29) when another group of subjects had to evaluate a single argument emphasizing the indicant: ‘‘The chance that a person with a positive test result has the disease must be very high because, when a person has the disease the test *always* gives a positive result and this guarantees that people with positive results have a high probability of having the disease.’’

A third group of 365 subjects was given *both* arguments to evaluate after the first administration of the problem (181 receiving the normative argument first and 184 receiving the normative argument second). As the third analysis in Table 9 indicates, after being presented with both arguments, significantly more subjects moved in the nonnormative (indicant) direction than in the normative (Bayesian) direction (34.9% versus 18.5%). Thus, when presented with arguments explicating reasons for both choices, subjects were somewhat more responsive to the argument for the nonnormative choice.

## GENERAL DISCUSSION

An application of Slovic and Tversky’s (1974) understanding/acceptance principle indicated that with respect to the sunk cost effect, Newcomb’s problem, and choosing P(D/∼H) and P(H) in the selection task, more reflective and engaged reasoners were more likely to respond in accord with the principles that define normative reasoning. Subjects giving the response considered normative on each of these tasks were either significantly higher in cognitive ability or in need for cognition than subjects giving the nonnormative response. In the case of the sunk cost problem, subjects giving the normative response scored significantly higher on both cognitive ability and need for cognition measures.

On these problems, when subjects were presented with both a normative and a nonnormative argument, in every case a significantly larger proportion of subjects changed from the nonnormative response to the normative response than changed from the normative response to the nonnormative alternative. Averaged across all four problems, 29.0% of the subjects changed in the normative direction, whereas only 11.7% changed in the nonnormative direction after evaluating arguments of both types.

This pattern was reinforced by a consideration of the conditions where only a single argument was presented. In three of four cases (Newcomb’s problem the exception) presenting a single nonnormative argument failed to yield significantly greater movement toward the nonnormative response than toward the normative response. In contrast, when a single normative argument was presented, in each of these four cases a significantly larger proportion of subjects moved in the normative direction than in the nonnormative direction. In this condition, averaged across the four problems, 42.0% of the subjects changed in the normative direction, whereas only 9.4% changed in the nonnormative direction. Finally, the combined effects of cognitive ability, need for cognition, and evaluating both arguments reduced the incidence of the nonnormative response by 32.6% in Newcomb's problem, 36.3% in the sunk cost problem, 17.5% for choosing P(D/∼H), and 30.4% for choosing P(H). According to the understanding/acceptance principle, all of these patterns are what we would expect if the standard normative model being applied to these tasks is the correct one to apply and if some people initially fail to give the normative response because of inadequate processing of the problem.

These patterns were, however, completely absent in the case of the prisoner's dilemma, cab problem, and disease problem. In none of these tasks did the group of subjects giving the nonnormative response have significantly lower cognitive ability or need for cognition scores. Moreover, in the condition where both arguments were presented, there was no significant tendency for responses to move disproportionately in the normative direction.

Two of the three problems failing to show the expected trends were noncausal base rate problems. Furthermore, the negative trends on these two problems were stronger than on the prisoner's dilemma. On the latter, several trends were in the direction of the other four tasks—just not significantly so. In contrast, on the noncausal base rate problems, sometimes the trends actually pointed significantly in the opposite direction (see Stanovich & West, 1998, for further anomalous findings involving noncausal base rates). For example, in the disease problem, a significantly larger proportion of subjects moved in the nonnormative direction after evaluating both arguments. Correspondingly, in the cab problem, the indicant group had SAT scores significantly higher than those of the Bayesian group.

These unexpected patterns interestingly mirror the history of contentious disputes surrounding these problems in the psychological and philosophical literature. The cab and the disease problem have spawned almost two decades of critical comment. The issues in dispute range from arguments that inappropriate normative models have been assumed (Birnbaum, 1983; Cohen, 1981, 1986; Gigerenzer, 1991, 1993; Kyburg, 1983; Levi, 1983, 1996) to arguments that these problems are linguistically and pragmatically unclear to subjects (Gigerenzer, Hell, & Blank, 1988; Hilton, 1995; Koehler, 1996; Macchi, 1995; Margolis, 1987; Schwarz, Strack, Hilton, & Naderer, 1991). Our results reinforce the concerns about the appropriateness of the normative models that have been applied to these tasks. Alternatively, the results may be indicating that these problems provoke alternative task construals that are equally rational[3] (see Stanovich, 1999). With respect to other tasks, how-

[3] A final possibility is that, with regard to these problems, there are true individual differences in rational thought (Stanovich, 1999; Stanovich & West, 1998)—that some people are

ever—Newcomb's problem, choosing P(D/∼H), the sunk cost problem—an application of the understanding/acceptance principle serves to reinforce the task construal intended by the experimenter and the normative analysis traditionally applied to the task.

Two caveats are in order at this point in the discussion. First, it might be argued that some of the differences obtained, although statistically significant, are quite modest in terms of effect size. For example, the two significant effects on SAT scores in Table 1 correspond to effect sizes of .229 (Cohen's d, see Rosenthal & Rosnow, 1991) and .204, respectively. The three significant effects on need for cognition scores in Table 1 correspond to effect sizes of .204, .222, and .193, respectively. The four task interpretations reinforced by the understanding/acceptance principle—those for choosing P(D/∼H), choosing P(H), the sunk cost problem, and Newcomb's problem—displayed phi coefficients in the two-argument condition of .16, .18, .34, and .21, respectively.

Nevertheless, despite these modest effects, it can be argued that the understanding/acceptance principle, as developed in the previous discussion, focuses attention on the *direction* of the association between response change and understanding and whether this direction is consistently observed. Thus, directionality and significance are the critical issues here, not effect size. Additionally, it seems impossible to know what magnitude of response change would be considered large in response to a single paragraph-long argument. Given the context of such a minimal manipulation, phi coefficients in the .20 to .35 range might seem as large as could be expected.

Nevertheless, the previous caveat does suggest another that reflects a more serious limitation in the Slovic/Tversky methodology used here. A major limitation of the technique is the obvious fact that there is no way to randomly sample populations of arguments in order to generate comparable arguments for each position. There is no way to determine a priori whether the arguments were equally compelling psychologically—that is, whether the effects are due to normativity or to differential persuasiveness. Fortunately, however, because subjects did rate the arguments, it is possible to arrange some post hoc comparisons that shed light on whether the overall persuasiveness of the normative and nonnormative arguments were reasonably comparable.

First, we focus only on the two conditions where the subjects were given a single argument. Second, because subjects rated the arguments after responding to the problem, it is necessary to equate the degree of congruence between the argument and response that the subject made. That is, it is necessary to compare how they rated the strength of the normative and nonnorma-

---

systematically computing a nonnormative rule and are not prevented from computing the normative one because of lack of cognitive ability or a low tendency toward reflective thought (Baron, 1991; Kahneman & Tversky, 1996; Shafir, 1993, 1994).

TABLE 10

Mean Ratings of the Normative and Nonnormative Arguments for Each Task for Subjects Whose Initial Responses Were Either Congruent or Incongruent with the Argument (Number of Subjects in Parentheses)

| | Nonnormative argument | Normative argument | *t* value |
|---|---|---|---|
| Selection Task, P(D/∼H) | | | |
| Congruent | 3.92 (50) | 4.27 (64) | 1.60 |
| Incongruent | 3.54 (59) | 3.49 (45) | −0.22 |
| Selection Task, P(H) | | | |
| Congruent | 3.39 (83) | 4.45 (40) | 3.88*** |
| Incongruent | 2.85 (26) | 4.07 (69) | 4.22*** |
| Sunk Cost Problem | | | |
| Congruent | 5.02 (45) | 5.03 (69) | 0.04 |
| Incongruent | 3.43 (63) | 3.89 (36) | 1.62 |
| Newcomb's Problem | | | |
| Congruent | 4.56 (71) | 4.59 (42) | 0.14 |
| Incongruent | 2.81 (42) | 3.97 (67) | 4.33*** |
| Prisoner's Dilemma | | | |
| Congruent | 4.67 (43) | 4.76 (72) | 0.63 |
| Incongruent | 3.10 (68) | 3.38 (37) | 1.07 |
| Cab Problem | | | |
| Congruent | 4.13 (38) | 3.69 (30) | −1.59 |
| Incongruent | 3.06 (32) | 3.65 (46) | 2.52* |
| Disease Problem | | | |
| Congruent | 4.38 (50) | 3.91 (42) | −1.73 |
| Incongruent | 2.93 (44) | 2.71 (48) | −0.78 |

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

tive argument when for both groups of subjects the argument was congruent with their response. Similarly, the normative and nonnormative argument can be compared when for both groups of subjects the argument was incongruent with their response. For example, the first two comparisons in Table 10 compare the two arguments concerning P(D/∼H) in the selection task. Among subjects who received arguments congruent with the response they had given, the mean rating of the normative argument (4.27) did not differ significantly from the mean rating given the nonnormative argument (3.92). The same was true for the rating comparison among subjects who received arguments incongruent with the response they had given. The rating given the normative argument (3.49) was very similar to that given the nonnormative argument (3.54).

Across all of the comparisons in the Table, 10 of the 14 displayed no significant difference between the normative and nonnormative arguments among subjects whose response congruence with the argument was equated.

Four of the 14 comparisons did however display significant differences in the direction of the normative argument. In Newcomb's problem and the cab problem only one of the two groups of subjects (the incongruent group) displayed a significant difference favoring the normative argument. In the case of the cab problem, this difference is of less concern because the difference goes in the other direction (favoring the nonnormative argument) for the the congruent group. In Newcomb's problem, the congruent group rated the arguments virtually identically.

There is really only one problem where the argument ratings indicate that the structure of the arguments was seriously biased and that is for the arguments regarding the choice of P(H)—where the normative argument was rated significantly higher than the nonnormative argument in both groups. The differential response to the arguments for this problem must be interpreted with this confound between normativity and argument persuasiveness in mind. Outside of the case of P(H) in the selection task, the arguments do not seem to have been seriously biased in one direction or another. Of course, extensive pretesting could be done in future applications of the method to insure the persuasive equivalence in advance. However, one further indication that the present set of arguments were reasonable was the overall level of the ratings given by the subjects who received arguments that were incongruent with their responses. As can seen in Table 10, the mean strength ratings given by these subjects were mostly in the range of 3.0 to 4.0 and clustered around 3.5—exactly the middle of the scale employed (3.5 is midway between ''somewhat weak'' and ''somewhat strong'' on our scale). That subjects whose response was completely incongruent with the argument did not rate the argument as extremely weak provides a final indication that all of the arguments had at least some force even for subjects disinclined toward their conclusion.

## REFLECTIVE EQUILIBRIUM AND RESPONSE VARIABILITY

One important issue that is raised by these experiments is the issue of the lability of the responses that are the subject of debate in the heuristics and biases literature. Particularly important are implications for critiques of the heuristics and biases literature provoked by findings that the majority of subjects depart from a particular normative model. As mentioned in the introduction, what invariably provokes these critiques—often by philosophers defending the notion of impeccable human rational competence—is the finding that the modal subject departs from the normative response. Trying to avoid the conclusion that human cognition is systematically irrational in some domain (see Stein, 1996), critics committed to the principle of virtually ideal rational competence (e.g., Cohen, 1981; Wetherick, 1995) argue either that experimenters are applying the wrong normative model to the task or that, alternatively, they may be applying the correct normative model to the problem as set, but they have presented the problem misleadingly so that the subject adopts a different problem construal and ends up providing a normatively appropriate answer to a *different* problem (Adler, 1984, 1991; Berkeley & Humphreys, 1982; Broome, 1990; Hilton, 1995; Schick, 1987; Schwarz, 1996).

This entire critique is misguided if the finding that spawned it—that typical human performance departs from a specified normative model—displays systematic lability. Consider, then, how labile were responses on some of these problems. Collapsed across all of the problems, when presented with an argument on each side of the question, an average of 23.4% altered their responses on a readministration of the task. When presented with a single normative argument, 40.5% of our nonnormative subjects switched to the normative response on a readministration of the task. Note that in our procedure subjects were *not* told that the argument was *correct*. They were simply told to evaluate the argument, and were free to rate it as very weak if it conflicted with their previous response. Similarly, subjects in the both-arguments condition were perfectly free to rate as weak the argument that conflicted with their response and to view the compatible argument as strong and thus as justification for persisting with their previous response. Nevertheless, one quarter of our participants changed responses after seeing one conflicting argument along with a compatible one.

Consider further how the modal response on these tasks can change as we impose criteria for reflectiveness and cognitive engagement. Of the 661 subjects who completed the first administration of the selection task, only 217 (32.8%) chose the base rate, P(H). Thus, failure to choose the base rate is the clear modal response on the first administration. If we wished to explain this deviation from Bayesian logic without positing a systematic deviation from rational principles, we might wish to criticize the experimenter's application of the normative model and/or the formulation of the problem. For example, we might be prone to start thinking about whether the experimenter has used an inappropriately broad reference class in defining the base rate (Kyburg, 1996; Levi, 1996) or perhaps whether the subjects' concept of probability is Baconian rather than Pascallian (Cohen, 1983). However, a much simpler assumption is that perhaps many of the subjects hadn't fully thought through the implications of the base rate. The latter assumption would seem to have the advantage of parsimony—particularly in light of the following empirical finding: after seeing one normative and one nonnormative argument, 52.8% of the subjects in the upper half of the SAT and need for cognition distributions chose the base rate! Subsequent to a very slight intervention, choosing the base rate becomes the modal response; and because choosing it coincides with the normative model assumed by the originators of this problem, there is now no need to assume error on their part (note that the critiques of the heuristics and biases literature tend to ignore all situations where the modal response *coincides* with the normative model being applied—experimenters apparently make no mistakes in these situations). A normative/descriptive gap that is disproportionately created by subjects with superficial understanding and/or low task engagement provides no warrant for amending the application of standard normative models.

Newcomb's problem yields the same reversal in the modal response after the three factors associated with greater reflectiveness are applied. Of the 662 subjects who completed the first administration of the problem, only 246 (37.2%) chose two boxes. Again, the nonnormative one-box choice was the modal response. However, after seeing one normative and one nonnormative argument, 53.8% of the subjects in the upper half of the SAT and need for cognition distributions chose two boxes. Among this more reflective group who had the advantage of argument explication, the two-box choice is the modal response.

In reply to Cohen's (1981) well-known critique of the heuristics and biases literature—surely the most often cited of such critiques—Jepson, Krantz, and Nisbett (1983) argue that ''Cohen postulates far too broad a communality in the reasoning processes of the 'untutored' adult'' (p. 495). Jepson et al., we argue, were right on the mark. We would only expand the phrase to read ''far too broad a communality and stability in the reasoning processes of the 'untutored' adult.'' Variation and instability in responses can be analyzed—in conjunction with tools such as the understanding/acceptance principle—in ways that shed light on debates about the reasons why descriptive and normative models of human reasoning sometimes do not coincide.

## REFERENCES

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

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

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

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axioms de l'ecole americaine. *Econometrica*, **21,** 503–546.

Arkes, H. R. (1996). The psychology of waste. *Journal of Behavioral Decision Making*, **9,** 213–224.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, **35,** 124–140.

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

Bar-Hillel, M. (1984). Representativeness and fallacies of probability judgment. *Acta Psychologica*, **55,** 91–107.

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

Baron, J. (1991). Some thinking is irrational. *Behavioral and Brain Sciences*, **14,** 486–487.

Baron, J. (1993). *Morality and rational choice*. Dordrecht: Kluwer.

Baron, J. (1994a). Nonconsequentialist decisions. *Behavioral and Brain Sciences*, **17,** 1–42.

Baron, J. (1994b). *Thinking and deciding (Second Edition)*. Cambridge, England: Cambridge University Press.

Baron, J. (1998). *Judgment Misguided: Intuition and error in public decision making*. New York: Oxford University Press.

Berkeley, D., & Humphreys, P. (1982). Structuring decision problems and the ''bias heuristic.'' *Acta Psychologica*, **50,** 201–252.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, **45,** 1185–1195.

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

Bornstein, B., & Chapman, G. (1995). Learning lessons from sunk costs. *Journal of Experimental Psychology: Applied*, **1,** 251–269.

Braine, M. D. S., Connell, J., Freitag, J., & O'Brien, D. P. (1990). Is the base rate fallacy an instance of asserting the consequent? In K. Gilhooly, M. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thinking* (Vol. 1, pp. 165–180). New York: Wiley.

Broome, J. (1990). Should a rational agent maximize expected utility? In K. S. Cook & M. Levi (Eds.), *The limits of rationality* (pp. 132–145). Chicago: University of Chicago Press.

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.

Campbell, R. (1985). Background for the uninitiated. In R. Campbell & L. Sowden (Eds.), *Paradoxes of rationality and cooperation: Prisoner's dilemma and Newcomb's problem* (pp. 3–41). Vancouver: University of British Columbia Press.

Campbell, R., & Sowden, L. (Eds.) (1985). *Paradoxes of rationality and cooperation: Prisoner's dilemma and Newcomb's problem*. Vancouver: University of British Columbia Press.

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

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

Cohen, L. J. (1979). On the psychology of prediction: Whose is the fallacy? *Cognition*, **7,** 385–407.

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

Cohen, L. J. (1982). Are people programmed to commit fallacies? Further thoughts about the interpretation of experimental data on probability judgment. *Journal for the Theory of Social Behavior*, **12,** 251–274.

Cohen, L. J. (1983). The controversy about irrationality. *Behavioral and Brain Sciences*, **6,** 510–517.

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

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

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance ethics. *The Journal of Philosophy*, **76,** 256–282.

Dawes, R. M. (1988). *Rational choice in an uncertain world*. San Diego: Harcourt Brace Jovanovich.

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

Dawes, R. M., Mirels, H., Gold, E., & Donahue, E. (1993). Equating inverse probabilities in implicit personality judgments. *Psychological Science*, **4,** 396–400.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory & Cognition*, **24,** 644–654.

Doherty, M. E., & Mynatt, C. (1990). Inattention to P(H) and to P(D/∼H): A converging operation. *Acta Psychologica*, **75,** 1–11.

Doherty, M. E., Mynatt, C., Tweney, R., & Schiavo, M. (1979). Pseudodiagnositicity. *Acta Psychologica*, **43,** 111–121.

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,** 191–194.

Eells, E. (1982). *Rational decision and causality*. Cambridge: Cambridge University Press.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, **85,** 395–416.

Elgin, C. Z. (1996). *Considered judgment*. Princeton, NJ: Princeton University Press.

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

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. London: Erlbaum.

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

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory & Cognition*, **12,** 402–410.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives*, **7,** 159–171.

Frisch, D. (1993). Reasons for framing effects. *Organizational Behavior and Human Decision Processes*, **54,** 399–429.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, **101,** 75–90.

Gibbard, A., & Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In C. Hooker, J. Leach, & E. McClennen (Eds.), *Foundations and applications of decision theory* (Vol. 1, pp. 125–162). Dordrecht: Reidel.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond ''heuristics and biases.'' *European Review of Social Psychology*, **2,** 83–115.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 284–313). London: Routledge.

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

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, **14,** 513–525.

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

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review,* **98,** 506–528.

Goodman, N. (1965). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Hargreaves Heap, S. P., & Varoufakis, Y. (1995). *Game theory: A critical introduction*. London: Routledge.

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

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

Hurley, S. L. (1991). Newcomb's problem, prisoner's dilemma, and collective action. *Synthese*, **86,** 173–196.

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

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47,** 263–291.

Kahneman, D., & Tversky, A. (1983). Can irrationality be intelligently discussed? *Behavioral and Brain Sciences*, **6,** 509–510.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, **39,** 341–350.

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

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

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

Kyburg, H. E. (1983). Rational belief. *Behavioral and Brain Sciences*, **6,** 231–273.

Kyburg, H. E. (1996). Probabilistic fallacies. *Behavioral and Brain Sciences*, **19,** 31.

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

Larrick, R. P., Smith, E. E., & Yates, J. F. (1992, November). *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.

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

Levi, I. (1996). Fallacy and controversy about base rates. *Behavioral and Brain Sciences*, **19,** 31–32.

Lewis, D. (1979). Prisoner's Dilemma is a Newcomb Problem. *Philosophy and Public Affairs*, **8,** 235—240.

Lopes, L. L., & Oden, G. C. (1991). The rationality of intelligence. In E. Eells & T. Maruszewski (Eds.), *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science* (pp. 199–223). Amsterdam: Editions Rodopi.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica*, **40,** 287–298.

Macchi, L. (1995). Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental Psychology*, **48A,** 188–207.

MacCrimmon, K. R. (1968). Descriptive and normative implications of the decision-theory postulates. In K. Borch & J. Mossin (Eds.), *Risk and uncertainty* (pp. 3–32). London: Macmillan.

MacCrimmon, K. R., & Larsson, S. (1979). Utility theory: Axioms versus 'paradoxes.' In M. Allais & O. Hagen (Eds.), *Expected utility hypotheses and the Allais paradox* (pp. 333–409). Dordrecht: Reidel.

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

March, J. G. (1988). Bounded rationality, ambiguity, and the engineering of choice. In D. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions* (pp. 33–57). Cambridge: Cambridge University Press.

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

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

Miller, P. M., & Fagley, N. S. (1991). The effects of framing, problem variations, and providing rationale on choice. *Personality and Social Psychology Bulletin*, **17,** 517–522.

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

Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Ed.), *Essays in honor of Carl G. Hempel* (pp. 114–146). Dordrecht: Reidel.

Nozick, R. (1993). *The nature of rationality*. Princeton, NJ: Princeton University Press.

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

Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds*. New York: Wiley.

Raiffa, H. (1985). Back from prospect theory to utility theory. In M. Grauer, M. Thompson, & A. Wierzbicki (Eds.), *Plural rationality and interactive processes* (pp. 100–113). Berlin: Springer-Verlag.

Rapoport, A., & Chammah, A. (1965). *Prisoner's dilemma*. Ann Arbor, MI: University of Michigan Press.

Rasmusen, E. (1989). *Games and information: An introduction to game theory*. Oxford: Blackwell.

Rawls, J. (1971). *A theory of justice*. Oxford: Oxford University Press.

Resnik, M. D. (1987). *Choices: An introduction to decision theory*. Minneapolis: University of Minnesota Press.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis.* (Second Edition). New York: McGraw-Hill.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schick, F. (1987). Rationality: A third dimension. *Economics and Philosophy*, **3,** 49–66.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahweh, NJ: Erlbaum.

Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of ''irrelevant'' information. *Social Cognition*, **9,** 67–84.

Shafir, E. (1993). Intuitions about rationality and cognition. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 260–283). London: Routledge.

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition*, **50,** 403–430.

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, **24,** 449–474.

Shafir, E., & Tversky, A. (1995). Decision making. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (Vol. 3, pp. 77–100). Cambridge, MA: MIT Press.

Sieck, W., & Yates, J. F. (1997). Exposition effects on decision making: Choice and confidence in choice. *Organizational Behavior and Human Decision Processes*, **70,** 207–219.

Skyrms, B. (1996). *The evolution of the social contract*. Cambridge: Cambridge University Press.

Slovic, P. (1995). The construction of preference. *American Psychologist*, **50,** 364–371.

Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science*, **19,** 368–373.

Slugoski, B. R., Shields, H. A., & Dawson, K. A. (1993). Relation of conditional reasoning to heuristic processing. *Personality and Social Psychology Bulletin*, **19,** 158–166.

Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making*, **9,** 283–290.

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

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

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford: Oxford University Press.

Takemura, K. (1992). Effect of decision time on framing of decision: A case of risky choice behavior. *Psychologia*, **35,** 180–185.

Takemura, K. (1993). The effect of decision frame and decision justification on risky choice. *Japanese Psychological Research*, **35,** 36–40.

Takemura, K. (1994). Influence of elaboration on the framing of decision. *Journal of Psychology*, **128,** 33–39.

Thagard, P., & Nisbett, R. E. (1983). Rationality and charity. *Philosophy of Science*, **50,** 250–267.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, **1,** 39–60.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge: Cambridge University Press.

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

Wetherick, N. E. (1971). Representativeness in a reasoning problem: A reply to Shapiro. *Bulletin of the British Psychological Society*, **24,** 213–214.

Wetherick, N. E. (1995). Reasoning and rationality: A critique of some experimental paradigms. *Theory and Psychology*, **5,** 429–448.

Wolfe, C. R. (1995). Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory account. *Journal of Behavioral Decision Making*, **8,** 85–108.

Accepted October 7, 1998