Memory & Cognition
1998, 26 (1), 161-179

# Who uses base rates and $P(D/{\sim}H)$? An analysis of individual differences

KEITH E. STANOVICH
University of Toronto, Toronto, Ontario, Canada

and

RICHARD F. WEST
James Madison University, Harrisonburg, Virginia

In two experiments, involving over 900 subjects, we examined the cognitive correlates of the tendency to view $P(D/{\sim}H)$ and base rate information as relevant to probability assessment. We found that individuals who viewed $P(D/{\sim}H)$ as relevant in a selection task and who used it to make the proper Bayesian adjustment in a probability assessment task scored higher on tests of cognitive ability and were better deductive and inductive reasoners. They were less biased by prior beliefs and more data-driven on a covariation assessment task. In contrast, individuals who thought that base rates were relevant did not display better reasoning skill or higher cognitive ability. Our results parallel disputes about the normative status of various components of the Bayesian formula in interesting ways. It is argued that patterns of covariance among reasoning tasks may have implications for inferences about what individuals are trying to optimize in a rational analysis (J. R. Anderson, 1990, 1991).

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 $P(D/{\sim}H)$—that is, 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). For example, Doherty and Mynatt (1990) used a simple selection paradigm, in which subjects were asked to imagine that they were doctors examining a patient with a red rash. The subjects were shown four cards with information on the back and were asked to choose which pieces of information they would need in order to determine whether the patient had the disease "Digirosa." The four pieces of information were the percentage of people with Digirosa, the percentage of people without Digirosa, the percentage of people with Digirosa who have a red rash, and the percentage of people without Digirosa who have a red rash. These pieces of information corresponded to the four terms in the formula above: $P(H)$, $P({\sim}H)$, $P(D/H)$, and $P(D/{\sim}H)$. Because $P(H)$ and $P({\sim}H)$ are complements, only three pieces of information are necessary to calculate the posterior probability. However, $P(D/{\sim}H)$ clearly must be selected, because it is a critical component of the likelihood ratio. Nevertheless, 48.8% of the individuals in their sample failed to select the $P(D/{\sim}H)$ card.

Similarly, Beyth-Marom and Fischhoff (1983) had subjects directly evaluate the relevance of the various components of the Bayesian formula in a hypothetical problem; they found that, across several different conditions, from 20% to 50% of their sample deemed $P(D/{\sim}H)$ to be irrelevant. Finally, in a variety of covariation detection experiments, subjects have been found to underweight components of information (e.g., cell D in the $2 \times 2$ design) that are necessary for the estimation of $P(D/{\sim}H)$ (Arkes & Harkness, 1983; Kao & Wasserman, 1993; Levin, Wasserman, & Kao, 1993; Schustack & Sternberg, 1981; Wasserman et al., 1990).

The other deviation from Bayesian reasoning that has been the subject of intense investigation is the tendency for individuals to ignore or underweight the prior probability, $P(H)$, which is presented as a base rate in many experiments (Bar-Hillel, 1980, 1984, 1990; Casscells, Schoenberger, & Graboys, 1978; Dawes, Mirels, Gold,

This research was supported by Grant No. 410-95-0315 from the Social Sciences and Humanities Research Council of Canada to K.E.S. and by a James Madison University Program Faculty Assistance Grant to R.F.W. The authors thank Penny Chiappe, Alexandra Gottardo, Walter Sa, Robin Sidhu, and Ron Stringer for their assistance in data coding. Requests for reprints should be sent to K. E. Stanovich, Department of Applied Psychology, Ontario Institute for Studies in Education, University of Toronto, 252 Bloor Street West, Toronto, ON, Canada M5S 1V6 (e-mail: kstanovich@oise.utoronto.ca).

—Accepted by previous editor, Geoffrey R. Loftus

& Donahue, 1993; Fischhoff & Bar-Hillel, 1984; Lyon & Slovic, 1976; Tversky & Kahneman, 1982). People have been found to give insufficient weight to this quantity in some paradigms and to deny its relevance in others. For example, in the Doherty and Mynatt (1990) selection task, 46.5% of the subjects failed to select the *P*(H) card. Performance was better in Beyth-Marom and Fischhoff's (1983) relevance judgment paradigm, but even there, across several experiments, 15%–25% of the subjects judged *P*(H) to be irrelevant. In a different paradigm, in which subjects were given only a base rate, Beyth-Marom and Fischhoff found that 25%–40% of the subjects failed to use it. Research based on still other paradigms that require the base rate to be amalgamated with a likelihood ratio has demonstrated that many subjects underweight the base rate (Hammerton, 1973; Macchi, 1995; Poulton, 1994; Wolfe, 1995).

In this study, we focus on one aspect of the results of probabilistic reasoning experiments that has been largely overlooked: individual differences. The research literature in this domain has largely ignored this aspect of performance; theoretical discussion has focused almost exclusively on the modal response given on the various tasks that have been the center of attention. As a result, debates about the normative appropriateness of particular responses on reasoning tasks (see, e.g., L. J. Cohen, 1981, 1982, 1986; Eells & Maruszewski, 1991; Gigerenzer, 1991, 1993; Gigerenzer & Murray, 1987; Koehler, 1996) have largely ignored potentially relevant information contained in the pattern of variability and covariance displayed across tasks. For example, discussions of the base rate "fallacy" or "base rate neglect" imply that it is a nearly universal characteristic of human cognition. However, even in base rate problems with the most complex and infelicitous wording, anywhere from 10% to 30% of subjects give a normatively appropriate response (Bar-Hillel, 1980; Lyon & Slovic, 1976; Macchi, 1995). Likewise, although 46.5% of Doherty and Mynatt's (1990) subjects failed to select the *P*(H) card, and 48.8% failed to select the *P*(D/~H) card, 11.4% of their sample made the normatively appropriate selection of the *P*(H), *P*(D/H), and *P*(D/~H) cards.

We will argue here that such variability poses some questions that are largely unaddressed in the critiques of the normative models used in the psychological literature. For example, the phenomenon of base rate neglect has been enormously controversial. Critics have argued that the phenomenon is due, in part, to subtle linguistic complexities in the wording of the questions and to methodological quirks in base rate experiments (Adler, 1984; Braine, Connell, Freitag, & O'Brien, 1990; Gigerenzer, Hell, & Blank, 1988; Koehler, 1996; Macchi, 1995; Macdonald, 1986; Poulton, 1994; Schwarz, Strack, Hilton, & Naderer, 1991). A more fundamental criticism is that the overall normative framework for base rate problems has been incorrect (Birnbaum, 1983; L. J. Cohen, 1979, 1981, 1982, 1986; Gigerenzer, 1991, 1993; Gigerenzer & Murray, 1987; Koehler, 1993b, 1996; Koehler & Shaviro, 1990; Kyburg, 1983; Levi, 1983; Macdonald, 1986; Schum, 1990). For example, Gigerenzer and Hoffrage (1995) have argued that traditional base rate problems have asked subjects to provide singular subjective probabilities (single-event likelihoods), when, in fact, people operate in a frequentistic mode—that is, they estimate the number of occurrences across a series of events (see also Cosmides & Tooby, 1996). In essence, Gigerenzer (1991, 1993; Gigerenzer, Hoffrage, & Kleinbolting, 1991) argues that subjects' performances have been evaluated against the wrong normative model in such tasks and that what has been termed the "base rate fallacy" is in actuality no fallacy at all. None of these critiques, however, provide an explanation of why an irreducible minority of subjects always gives the response that was considered correct by the experimenters who designed the problem.

Although the normative correctness of evaluating *P*(D/~H) is much less controversial than is the use of base rates, investigators have questioned the focus on the complement hypothesis that is stressed by the traditional falsificationist research strategy. Klayman and Ha (1987, 1989) have illustrated how certain task environments make a hypothesis-testing strategy that concentrates on the focal hypothesis quite efficacious. Similarly, Friedrich (1993) argues that, if the human cognitive apparatus is designed to avoid certain types of predictive errors rather than to seek the truth (see Einhorn & Hogarth, 1978; Halberstadt & Kareev, 1995; Stein, 1996; Stich, 1990), then focusing on the focal hypothesis and showing relative inattention to its complement might be a processing pattern that is to be expected (see also McKenzie, 1994).

Friedrich's (1993) argument is in the tradition of optimization models (Schoemaker, 1991) that emphasize the adaptiveness of human cognition (J. R. Anderson, 1990, 1991; Campbell, 1987; Cosmides & Tooby, 1994, 1996; Oaksford & Chater, 1993, 1994, 1995). For example, J. R. Anderson (1990, 1991)—building on the work of Marr (1982), Newell (1982), and Pylyshyn (1984)—defines four levels of theorizing in cognitive science: a biological level that is inaccessible to cognitive theorizing, an implementation level designed to approximate the biological, an algorithmic level (an abstract specification of the steps necessary to carry out a process), and the rational level. The last level provides a specification of the goals of the system's computations (*what* the system is attempting to compute and *why*) and can suggest constraints on the operation of the algorithmic level. According to J. R. Anderson (1990), the rational level specifies what are the "constraints on the behavior of the system in order for that behavior to be optimal" (p. 22). The description of this level of analysis proceeds from a "general principle of rationality" which assumes that "the cognitive system operates at all times to optimize the adaptation of the behavior of the organism" (p. 28). Thus, the rational level of analysis is concerned with the goals of the system, with beliefs relevant to those goals, and with the choice of action that is rational, given the system's goals and beliefs (J. R. Anderson, 1990; Bratman, Israel, & Pollack, 1991; Dennett, 1987; Newell, 1982, 1990).

However, even if humans are optimally adapted to their environments at the rational level of analysis, there may still be computational limitations at the algorithmic level that prevent the full realization of the optimal model (see, e.g., Cherniak, 1986; Goldman, 1978; Oaksford & Chater, 1993, 1995). Even if we assume that the rational model for all humans in a given environment is the same, we would still expect there to be individual differences in actual performance (despite no rational level differences) because of differences at the algorithmic level. We would assume that the responses of organisms with fewer algorithmic limitations would be closer to the response that a rational analysis would reveal as optimal. Thus, the direction of the correlation between response type and cognitive capacity provides an empirical clue about the nature of the optimizing response.

Alternatively, the direction of the correlation might still have implications even if we do not wish to make the assumption that the model being optimized is normative or rational (see, e.g., Baron, 1991b; Nisbett & Ross, 1980; Shafir, 1991, 1994; Stanovich, in press). In fact, J. R. Anderson (1990) himself accepts Stich's (1990) argument that evolutionary adaptation does not guarantee perfect human rationality in the normative sense: "Rationality in the adaptive sense, which is used here, is not rationality in the normative sense that is used in studies of decision making and social judgment.... It is possible that humans are rational in the adaptive sense in the domains of cognition studied here but not in decision making and social judgment" (p. 31). Thus, although Anderson's rational analysis proceeds from the central assumption that cognition is optimally adapted in an *evolutionary* sense, in most work in the judgment and decision literature, normative appropriateness, *not* optimal fitness in the evolutionary sense, is the prime concern.

Might the direction of the correlation between cognitive ability and response choice in a probabilistic reasoning task still tell us something about rationality under the normative, rather than adaptationist, view? The answer is yes if it is believed that descriptive facts about human cognition can be used as an inferential tool in deciding what is normative—and there are strong traditions in philosophy and psychology supporting such a belief. For example, some philosophers (e.g., L. J. Cohen, 1981, 1982) view untutored intuition as the sine qua non of normative justification. Stein (1996) notes that proponents of this position believe that the normative can simply be "read off" from a descriptive model because "whatever human reasoning competence turns out to be, the principles embodied in it are the normative principles of reasoning" (p. 231). In less extreme form, other theorists have argued that descriptive models of human behavior can, at least in part, condition our inferences about what is normative. March (1988) refers to this tradition when he discusses how actual human behavior has conditioned models of efficient problem solving in the areas of artificial intelligence and of organizational decision making. Likewise, Slovic (1995) refers to the "deep interplay between descriptive phenomena and normative principles" (p. 370). Thagard and Nisbett (1983) argue that the "discovery of discrepancies between inferential behavior and normative standards may in some cases signal a need for revision of the normative standards, and the descriptions of behavior may be directly relevant to what revisions are made" (p. 265; see also Kornblith, 1985, 1993; Kyburg, 1983; Shafer, 1988; Thagard, 1982).

Theorists who make the argument in favor of taking the inductive leap from the descriptive to the normative almost always take the modal response as the descriptive aspect of behavior that they wish to project. But we must ask whether the modal response is the *only* aspect of group performance that is relevant. Might the pattern of responses *around* the mode tell us something? And finally, what about the rich covariance patterns that would be present in any multivariate experiment? If we are to infer something about the normative from the descriptive (and of course not all investigators are agreed that we should), the thesis to be explored here is that there is more information available for such an inference than has traditionally been relied upon. Larrick, Nisbett, and Morgan (1993) made such an argument in their analysis of what justified the cost–benefit reasoning of microeconomics: "Intelligent people would be more likely to use cost–benefit reasoning. Because intelligence is generally regarded as being the set of psychological properties that makes for effectiveness across environments ... intelligent people should be more likely to use the most effective reasoning strategies than should less intelligent people" (p. 333). Larrick et al. are alluding to the fact that we may want to condition our inferences about the normative, not only on the basis of what response the majority of people make, but also on the basis of what response the most cognitively competent subjects make. If the normative response is not only more efficacious but also more computationally complex, we might expect that it would only be computed by those subjects with greater cognitive capacity. Alternatively, the normative strategy might *not* be more computationally complex. It might simply be more efficient and more readily recognized as such by individuals who are more intelligent. Either way, Larrick et al.'s suggestion holds—if we do want to condition normative models on the basis of descriptive models, the direction of the correlation with intellectual resources might provide useful information.

Another way to think about this argument is in terms of the positive manifold present in virtually all groups of cognitive tasks—the fact that different measures of cognitive ability almost always correlate with each other (see Carroll, 1993). The argument is that the empirical fact of a positive manifold can be put to use in those areas of cognitive psychology where the nature of the normative response is in dispute. The point is that scoring a vocabulary item on a cognitive ability test and scoring a probabilistic reasoning response on a task from the heuristics and biases literature are not the same. The correct response in the former task has a canonical interpretation agreed upon by all investigators, whereas the normative appropriateness of responses on tasks from the latter domain has been the subject of extremely contentious disputes (Berkeley & Humphreys, 1982; Birnbaum, 1983; L. J. Cohen, 1981, 1982, 1986; Cosmides & Tooby, 1996; Einhorn & Hogarth, 1981; Gigerenzer, 1991, 1993, 1996; Hilton, 1995; Kahneman & Tversky, 1996; Koehler, 1996; Macchi, 1995; Nickerson, 1996; Stein, 1996). A positive manifold between the two classes of task would only be expected if the normative model being used for directional scoring of the tasks in the latter domain is correct. Likewise, given that a positive manifold is the norm among cognitive tasks, the lack of a correlation (or a negative correlation) between a probabilistic reasoning task and more standard cognitive ability measures might be taken as a signal that the wrong normative model is being applied to the former task or that there are alternative models that are normatively appropriate.

In short, debates about which responses are optimal, normative, or prescriptive (see Baron, 1985; Bell, Raiffa, & Tversky, 1988; Evans, Over, & Manktelow, 1993) on probabilistic reasoning tasks might be leavened by a more detailed knowledge of just *who* was making which response and of how these people responded on other indicators of cognitive ability and reasoning skill (Stanovich, in press). For example, theorists in the heuristics and biases literature who defend the standard normative models are sometimes criticized for explaining divergences between normative models and actual performance by claiming that limitations in computational capacity prevent the normative response. But critics who claim that the wrong normative model is being invoked have argued that there is "no support for the view that people would choose in accord with normative prescriptions if they were provided with increased capacity" (Lopes & Oden, 1991, p. 209). One way to indirectly test this claim is to investigate how responses on Bayesian probability judgment tasks correlate with measures of cognitive capacity.

In the present study, we employed this strategy by examining whether individuals who judge base rates as relevant and who pay attention to $P(D/{\sim}H)$ when evaluating evidence exceed those who do not in cognitive ability and whether they reason better on other well-known deductive and inductive reasoning tasks. In the following two experiments, cognitive capacity was operationalized by well-known cognitive ability and academic aptitude tasks. All are known to load highly on psychometric *g* (Carroll, 1993; Matarazzo, 1972), and such measures have been linked to neurophysiological and information-processing indicators of efficient cognitive computation (Caryl, 1994; Deary, 1995; Deary & Stough, 1996; Detterman, 1994; Fry & Hale, 1996; Hunt, 1987; Stankov & Dunn, 1993; Vernon, 1991, 1993; Vernon & Mori, 1992). The psychometric and information-processing characteristics of the inductive and deductive reasoning tasks are less fully worked out, but they have repeatedly been viewed as prime exemplars of reasoning ability (Evans, Newstead, & Byrne, 1993; Johnson-Laird & Byrne, 1991; Nisbett, 1993).

## EXPERIMENT 1

### Method

#### Subjects

The subjects were 360 undergraduate students (138 males and 222 females) recruited through an introductory psychology subject pool at a medium-sized state university. Their mean age was 18.9 years ($SD = 2.2$). The demographics form filled out by the students included questions on their educational history in mathematics and statistics courses. We constructed a 0–4 point scale that assessed the student's mathematics/statistics course background. Students received 1 point if they had taken a statistics course in college (94 students), 1 point if they had taken a statistics course in high school (46 students), 1 point if they had taken a mathematics course in college (280 students), and 1 point if they had had 4 years of high school mathematics (314 students). The mean score on the scale was 2.04 ($SD = 0.82$). The subjects were also asked if they had taken a logic course in college or high school. Because only a few subjects had had a logic course in high school (31) or college (27), we constructed a 0/1 variable which was scored 1 if the subject had taken a logic course in high school or college.

#### Bayesian Reasoning Tasks

**Information selection task.** The information selection task was a slight variant of the task used by Doherty and Mynatt (1990). Subjects were given the following instructions:

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

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

**Probability assessment task.** These two-step problems were adapted from problems used in Experiment 5 of Beyth-Marom and Fischhoff (1983). The instructions for the first problem, hereafter termed the *David (.25, .70/.90) problem*, were as follows:

> Imagine yourself meeting David Maxwell. Your task is to assess the *probability that he is a university professor* based on some information that you will be given. This will be done in two steps. At each step you will get some information that you may or may not find useful in making your assessment. After each piece of information you will be asked to assess the probability that David Maxwell is a university professor. In doing so, consider all the information you have received to that point if you consider it to be relevant. Your probability assessments should be numbers between 0 and 1 that express your degree of belief. 1 means "I am absolutely certain that he is a university professor." 0 means "I am absolutely certain he is not a university professor." .65 means "The chances are 65 out of 100 that he is a university professor," and so forth. You can use any number between 0 and 1, for example, .15, .95, etc.
>
> Step One: You are told that David Maxwell attended a party in which 25 male university professors and 75 male business executives took part, 100 people all together. Question: What do you think the probability is that David Maxwell is a university professor? ____
>
> Step Two: You are told that David Maxwell is a member of the Bears Club. 70% of the male university professors at the above-mentioned

> party were members of the Bears Club, and 90% of the male business executives at the party were members of the Bears Club. Question: What do you think the probability is that David Maxwell is a university professor? ____

Thus, in this problem, reliance on the base rate at Step 1 would result in an estimate of .25. Step 2 is structured so that, although the likelihood ratio is less than 1 (.70/.90), P(D/H) is greater than .50. This might suggest to someone who ignored P(D/~H)—which is in fact higher than P(D/H)—that they should increase the probability that David is a university professor. Conversely, because the proper Bayesian adjustment is from .25 in Step 1 to .206 in Step 2—$(.70 \times .25)/(.70 \times .25 + .90 \times .75)$—any adjustment downward from Step 1 to Step 2 would suggest that the subject had been attentive to P(D/~H).

The second problem, hereafter termed the *Mark (.80, .40/.05) problem*, was phrased as follows:

> Again, imagine yourself meeting Mark Smith. Your task is to assess the *probability that he is a university professor* based on some information that you will be given.
>
> Step One: You are told that Mark Smith attended a party in which 80 male university professors and 20 male business executives took part, 100 people all together. Question: What do you think the probability is that Mark Smith is a university professor? ____
>
> Step Two: You are told that Mark Smith is a member of the Bears Club. 40% of the male university professors at the above mentioned party were members of the Bears Club, and 5% of the male business executives at the party were members of the Bears Club. Question: What do you think the probability is that Mark Smith is a university professor? ____

Thus, in this problem, reliance on the base rate at Step 1 would result in an estimate of .80. Step 2 is structured so that, although the likelihood ratio is considerably greater than 1 (.40/.05), P(D/H) is less than .50. This might suggest to someone who ignored P(D/~H)—which is in fact lower than P(D/H)—that these data should *decrease* the probability that David is a university professor. Conversely, because the proper Bayesian adjustment is from .80 in step one to .97 in Step 2—$(.40 \times .80)/(.40 \times .80 + .05 \times .20)$—any adjustment upward from Step 1 to Step 2 would suggest that the subject had been attentive to P(D/~H).

### General Ability Measures

**Scholastic Aptitude Test scores.** Because Scholastic Aptitude Test (SAT) scores were not available to us because of university restrictions, students were asked to indicate their verbal and mathematical SAT scores on a demographics sheet. The mean reported verbal SAT score (SAT–V) of the 349 students who filled in this part of the questionnaire was 529 ($SD = 72$), the mean reported mathematical SAT score (SAT–M) was 578 ($SD = 72$), and the mean total SAT score was 1,107 ($SD = 108$). These reported scores match the averages for this institution (520, 587, and 1,107) quite closely (Straughn & Straughn, 1995). A further indication of the validity of the self-reported scores is that the correlation (.49) between a vocabulary test (described below) and the reported SAT total score was quite similar to the .51 correlation between the vocabulary checklist and *verified* total SAT scores in a previous investigation that used the same vocabulary measure (West & Stanovich, 1991). A final indication of the validity of the SAT reports is that the vocabulary test displayed a higher correlation (.61) with the verbal SAT scores than with the mathematical SAT scores (.13). The difference between these dependent correlations (see J. Cohen & P. Cohen, 1983, pp. 56–57) was highly significant ($p < .001$).

**Vocabulary test.** As an additional converging measure of cognitive ability to supplement the SAT scores, a brief vocabulary measure was administered to the subjects (because vocabulary is the strongest specific correlate of general intelligence; see Matarazzo, 1972). This task employed the checklist-with-foils format that has been shown to be a reliable and valid way of assessing individual differences in vocabulary knowledge (R. C. Anderson & Freebody, 1983; Cooksey & Freebody, 1987; White, Slater, & Graves, 1989; Zimmerman, Broder, Shaughnessy, & Underwood, 1977). The stimuli for the task were 40 words and 20 pronounceable nonwords taken largely from the stimulus list of Zimmerman et al. (1977). The words and nonwords were intermixed through alphabetization. The subjects were told that some of the letter strings were actual words whereas others were not and that their task was to read through the list of items and to put a check mark next to those that they knew were words. Scoring on the task was determined by taking the proportion of the target items that were checked and subtracting the proportion of foils checked. Other corrections for guessing and differential criterion effects (see Snodgrass & Corwin, 1988) produced virtually identical correlational results.

### Deductive and Inductive Reasoning Tasks

**Syllogistic reasoning.** Twenty-four syllogistic reasoning problems, largely drawn from Markovits and Nantel (1989), were completed by the subjects. Eight of the problems were worded such that the validity judgment was in conflict with the believability of the conclusion (e.g., All flowers have petals; Roses have petals; therefore, Roses are flowers—which is invalid). Eight of the problems were worded such that the validity judgment was congruent with the believability of the conclusion (e.g., All fish can swim; Tuna are fish; therefore, Tuna can swim—which is valid). Eight of the problems involved imaginary content (e.g., All opprobines run on electricity; Jamtops run on electricity; therefore, Jamtops are opprobines—which is invalid).

Subjects were instructed as follows: "In the following problems, you will be given two premises *which you must assume are true.* A conclusion from the premises then follows. You must decide whether the conclusion *follows logically* from the premises or not. You must *suppose that the premises are all true* and limit yourself only to the information contained in the premises. This is very important. Decide if the conclusion follows logically from the premises, assuming the premises are true, and circle your response." After each item, the subjects indicated their responses by circling one of the two alternatives: a. Follows Logically, b. Does Not Follow Logically.

Although, as in previous experiments (Markovits & Nantel, 1989; Newstead, Pollard, Evans, & Allen, 1992), subjects performed better on problems where the believability of the conclusion was congruent with logical validity, the correlations with other variables were the same for all three types of syllogistic reasoning problems. Thus, the total number of correct responses across all 24 problems will be used in the analyses that follow. The mean score was 18.9 ($SD = 4.1$).

**Statistical reasoning.** Six problems were adapted from the work of Fong, Krantz, and Nisbett (1986) and Jepson, Krantz, and Nisbett (1983) and were structured so that the subject had to make an inductive inference in a simulation of a real-life situation. The information relevant to the decision was conflicting and of two different types. One type of evidence was statistical—either probabilistic or aggregate base rate information that favored one of the bipolar decisions. The other evidence was a concrete case, a singular instance, or a personal experience that pointed in the opposite direction. An example of the six items is the well-known "Volvo problem" (see p. 285 of Fong et al., 1986):

> The Caldwells had long ago decided that when it was time to replace their car they would get what they called "one of those solid, safety-conscious, built-to-last Swedish" cars—either a Volvo or a Saab. When the time to buy came, the Caldwells found that both Volvos and Saabs were expensive, but they decided to stick with their decision and to do some research on whether to buy a Volvo or a Saab. They got a copy of Consumer Reports and there they found that the consensus of the experts was that both cars were very sound mechanically, although the

Volvo was felt to be slightly superior on some dimensions. They also found that the readers of Consumer Reports who owned a Volvo reported having somewhat fewer mechanical problems than owners of Saabs. They were about to go and strike a bargain with the Volvo dealer when Mr. Caldwell remembered that they had two friends who owned a Saab and one who owned a Volvo. Mr. Caldwell called up the friends. Both Saab owners reported having had a few mechanical problems but nothing major. The Volvo owner exploded when asked how he liked his car. "First that fancy fuel injection computer thing went out: $400 bucks. Next I started having trouble with the rear end. Had to replace it. Then the transmission and the clutch. I finally sold it after 3 years at a big loss." What do you think the Caldwells should do?

The problem was followed by the choices: (1) They should definitely buy the Saab. (2) They should probably buy the Saab. (3) They should probably buy the Volvo. (4) They should definitely buy the Volvo. A preference for the Volvo would indicate a tendency to rely on the large-sample information in spite of the salient personal testimony. A preference for the Saab indicates reliance on the personal testimony over the opinion of experts and the large-sample information. Five additional problems of this type were employed: the college choice, admissions, and class choice problems adapted from Jepson et al. (1983), and the curriculum choice and marriage/baseball performance problems adapted from Fong et al. (1986). The problems were all scored in the direction giving higher scores to the choice made on the basis of the aggregate information and lower scores to choice made on the basis of the singular evidence. Performance on each of the six statistical reasoning items was standardized and the six $z$ scores were summed to form a composite score.

**Procedure**

The subjects completed the tasks during a single 2-h session in which they also completed some other tasks not part of the present investigation. They were tested in small groups of 3 to 8 individuals. Because a few subjects left certain forms blank and because of printing/collating errors in some of the materials, fewer than 360 subjects completed some of the tasks. Specifically, 352 subjects completed the information selection task, 347 subjects the David (.25, .70/.90) problem, 349 subjects the Mark (.80, .40/.05) problem, and 356 subjects the syllogistic reasoning problems.

## Results

### Information Selection Task

The choices observed on our version of the information selection task were roughly convergent with those observed by Doherty and Mynatt (1990). Four patterns accounted for over 93% of the choices in our study. The normatively correct choice of $P$(H), $P$(D/H), and $P$(D/~H) was made by 13.4% of our sample, compared to 11.4% of theirs. The most popular choice in our sample was for the two components of the likelihood ratio, $P$(D/H) and $P$(D/~H). Choosing these two cards only was the response of 35.5% of our sample (and of 29.5% of the Doherty & Mynatt sample). More subjects in our study (21.9%) chose $P$(D/H) only than they did in their study (12.3%). The choice of the base rate, $P$(H), and the numerator of the likelihood ratio, $P$(D/H)—while ignoring the denominator of the likelihood ratio, $P$(D/~H)—was somewhat less popular in our study (22.7%) than it was in theirs (33.6%). These differences, which we replicate in Experiment 2, might be due to differences in the physical layout of the alternatives between our experiments and Doherty and Mynatt's study. In our studies, $P$(H) and $P$(D/H) were the second and fourth sentences in a vertical list. In the Doherty and Mynatt study, the four alternative sentences were presented within squares representing "cards" in a 2 × 2 layout. The cards representing $P$(H) and $P$(D/H) were the two leftmost cards, with $P$(H) being directly above $P$(D/H). This layout might have encouraged more linkage of those alternatives in their study.

In the present study, almost all of the subjects (96.0%) viewed the $P$(D/H) card as relevant and very few (2.8%) viewed the $P$(~H) card as relevant. Overall, 54.3% of our subjects chose $P$(D/~H) as a necessary card, and 41.5% of the sample thought it was necessary to know the base rate, $P$(H). Thus, approximately half of the sample thought that the base rate was irrelevant, and about half thought that $P$(D/~H) was irrelevant. And, of course, these were not the same people. Less than 22% of the sample viewed *both* the base rate and $P$(D/~H) as irrelevant.

The SAT scores of the subjects who did and of those who did not include $P$(D/~H) in their choices are compared on the left side of Figure 1. The mean SAT–M score of subjects who selected the $P$(D/~H) card as one of their choices was significantly higher than that of individuals not choosing that card [585 vs. 568, $t(339) = 2.18$, $p < .05$; 95% confidence interval for the difference = 1.6 to 32.1]. The mean SAT–V score was likewise significantly higher [537 vs. 520, $t(339) = 2.22$, $p < .05$; 95% confidence interval for the difference = 1.9 to 32.4]. Similarly, a converging measure of general ability—the vocabulary test—displayed a significant difference [.580 vs. .536, $t(350) = 2.52$, $p < .025$; 95% confidence interval for the difference = .010 to .077] favoring those who chose $P$(D/~H). There were no significant differences on the syllogistic reasoning or statistical reasoning tasks, although the differences were in the same direction. Choos-

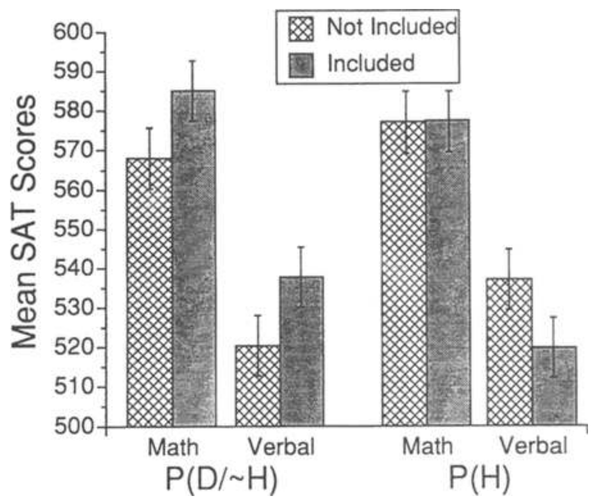


**Figure 1. Mean SAT scores as a function of whether the subjects included or did not include $P$(D/~H) and $P$(H) in their choices on the selection task of Experiment 1. Error bars represent the 95% confidence intervals for the difference between the pairs of sample means. In this figure, the error bars span a mean by a distance equal to the critical value of $t$ multiplied by the estimated standard error of the difference between sample means.**

ing $P(\text{D}/{\sim}\text{H})$ was not associated with educational history. There was no significant difference in mathematics background as assessed by the mathematics background composite variable. Few subjects had had a logic course in either college or high school, but those who had were no more likely to choose $P(\text{D}/{\sim}\text{H})$. In fact, the proportion of the sample who chose $P(\text{D}/{\sim}\text{H})$ and who had taken such a course (11.0%) was lower than the proportion who did not choose $P(\text{D}/{\sim}\text{H})$ and who had taken such a course (20.5%).

The SAT scores of the subjects who did and of those who did not include $P(\text{H})$ in their choices are compared on the right side of Figure 1. The results were quite different from those obtained when the sample was split on the basis of $P(\text{D}/{\sim}\text{H})$. The subjects who selected the base rate as one of their choices did not attain higher scores on the tests of general ability and reasoning. In fact, there were mild trends in the opposite direction. Although there was no difference on the SAT–M, there was a significant difference favoring the group *not* choosing the base rate on the verbal section of the SAT [537 vs. 520, $t(339) = -2.21$, $p < .05$; 95% confidence interval for the difference = 1.9 to 32.4]. There were no significant differences on the vocabulary test or the syllogistic reasoning task. Interestingly, there was a significant difference in performance on the statistical reasoning task, and the difference was in an unexpected direction—subjects choosing the base rate on the information selection task were *less* likely to rely on aggregate information in the statistical reasoning problems [$-.334$ vs. .324, $t(350) = -2.14$, $p < .05$; 95% confidence interval for the difference = .053 to 1.262]. If the aggregate information is considered to be conceptually similar to the base rate in the selection task, then one would have expected that base rate choosers would score higher on the statistical reasoning measure, which is scored in the direction of the aggregate choice (the direction of this relationship remains the same in Experiment 2, but it does not come close to attaining statistical significance). Finally, there were no significant differences between the 46 individuals who made the normatively appropriate choice—[$P(\text{H})$, $P(\text{D}/\text{H})$, and $P(\text{D}/{\sim}\text{H})$]—and the rest of the sample on any of the cognitive variables, primarily because of the countervailing trends associated with choosing $P(\text{H})$ and $P(\text{D}/{\sim}\text{H})$. There was a small but statistically significant negative association between the choice of $P(\text{H})$ and $P(\text{D}/{\sim}\text{H})$; 62.6% of the subjects who did not choose $P(\text{H})$ did choose $P(\text{D}/{\sim}\text{H})$, whereas only 42.5% of the subjects who chose $P(\text{H})$ also chose $P(\text{D}/{\sim}\text{H})$ [$\chi^2(1) = 13.19$, $p < .001$].

Taken collectively, the results indicate that people who picked $P(\text{D}/{\sim}\text{H})$ were better reasoners on other independent tests, but people who picked the base rate were not. The latter finding held at the extremes of the distribution as well. Of the 35 individuals (roughly 10%) with the lowest SAT scores (mean SAT total of 903), 42.9% chose $P(\text{H})$—a percentage very similar to that in the sample as a whole (41.5%). Only 31.4% of the 35 individuals with the highest SAT scores (mean SAT total of 1,288) chose $P(\text{H})$—a percentage somewhat lower than that in the sample as a whole.

**Probability Assessment Task**

In Step 1 of the David (.25, .70/.90) problem, 84.7% of the sample responded with the base rate of .25. This is somewhat higher than the 65% to 75% who responded with the base rate across the various conditions of the Beyth-Marom and Fischhoff (1983) experiment. Thus, many more subjects, when presented with *only* a base rate, relied on it for their judgments than deemed it relevant in the information selection task (84.7% vs. 41.5%). In the latter task, the subject must choose the base rate when it is presented along with other useful information (the components of the likelihood ratio). In the probability assessment task, it is the only information available at that step. Thus, the pragmatic cues of the experiment (see Hilton, 1995; Levinson, 1995) suggest its usefulness in the probability assessment task.

Subjects failing to respond with the base rate at Step 1 were not generally governed by the principle of indifference (Keynes, 1921). Only 18.9% of the subjects not responding with .25 at Step 1 gave .50 as a response. *Unlike* the subjects who did not choose $P(\text{H})$ in the information selection task, subjects not responding with the base rate at Step 1 of the David (.25, .70/.90) problem *did* have lower SAT scores than those who gave .25 as a response [mean total SAT of 1,070 vs. 1,113, $t(336) = 2.66$, $p < .01$; 95% confidence interval for the difference = 11.1 to 74.6]. The base rate in this problem is disproportionately ignored by the subjects with less cognitive ability, whereas this was not the case in the selection task.

Performance on the Mark (.80, .40/.05) problem paralleled that on the David (.25, .70/.90) problem in that only 18.1% of the subjects did not respond with the base rate (.80) on Step 1. Again, subjects were not generally governed by the principle of indifference, because only 14.9% of the subjects not responding with the base rate at Step 1 gave .50 as a response. As in the David (.25, .70/.90) problem, subjects not responding with the base rate at Step 1 of the Mark (.80, .40/.05) problem had lower SAT total scores than those who did respond with .80 [1,068 vs. 1,116, $t(338) = 3.18$, $p < .01$; 95% confidence interval for the difference = 18.3 to 77.5]. As previously mentioned, pragmatic cues are present in the probability assessment task that suggest the appropriateness of choosing the base rate, whereas no such cues are present in the selection task. Because no differences in cognitive ability were observed in the latter task between those who chose $P(\text{H})$ and those who did not—and because so many more subjects responded with the base rate in the probability assessment task—it may be that the SAT difference on the probability assessment task is the result of differences in sensitivity to the pragmatic cues of relevance (Adler, 1984; Bless, Strack, & Schwarz, 1993; Hilton, 1995; Sperber, Cara, & Girotto, 1995) rather than of differences in probabilistic reasoning itself.

On the second step of the David (.25, .70/.90) problem, 146 subjects (42.1%) adjusted their probabilities downward (the normatively appropriate response with a likelihood ratio less than one) and 131 (37.8%) of the subjects adjusted their probabilities upward. The latter response would occur if an individual ignored $P(D/{\sim}H)$ and viewed the $P(D/H)$ value of .70 as positively diagnostic because it was greater than .50. A substantial minority of the subjects (70) did not change their probability from Step 1 to Step 2 after the presentation of the likelihood ratio. This response might have been the result of ignoring *both* components of the likelihood ratio. Alternatively, it might have been the result of subjects tempering their tendency to respond to a high $P(D/H)$ value because they noticed that $P(D/{\sim}H)$ was greater than $P(D/H)$. (Also, note that the correct Bayesian probability revision of .206 is not that far below .25.) At least it is clear that these subjects did not ignore $P(D/{\sim}H)$ and revise upward on the basis of a $P(D/H)$ of .70. Thus, although not normative, the no-change response seems to be closer to the normatively appropriate response of decreasing the probability than to the inappropriate response of increasing the probability.

On the second step of the Mark (.80, .40/.05) problem, 105 subjects (30.1%) adjusted their probabilities upward (the normatively appropriate response with a likelihood ratio greater than one) and 182 (52.1%) of the subjects adjusted their probabilities downward. The latter response would occur if an individual ignored $P(D/{\sim}H)$ and viewed the $P(D/H)$ value of .40 as negatively diagnostic because it is less than .50. A substantial minority of the subjects (62) did not change their probability from Step 1 to Step 2 after the presentation of the likelihood ratio. Again, this response might have been the result of ignoring *both* components of the likelihood ratio. Alternatively, it might have been the result of subjects tempering their tendency to respond to a low $P(D/H)$ value because they noticed that $P(D/{\sim}H)$ was lower than $P(D/H)$. It at least is certain that these subjects did not ignore $P(D/{\sim}H)$ and revise downward on the basis of a $P(D/H)$ of .40.

Across the David (.25, .70/.90) and Mark (.80, .40/.05) problems, 51.0% of the sample responded consistently to the likelihood ratio presented in Step 2: 52 subjects (termed the *normative* group) made a normatively appropriate probability adjustment both times, 43 subjects (termed the *no change* group) did not change their probabilities in either problem, and 82 subjects (termed the *nonnormative* group) adjusted their probabilities in the wrong direction on both problems [strongly suggesting that they were consistently ignoring $P(D/{\sim}H)$]. Table 1 presents the means of these three groups of subjects on the other tasks in the study. There were significant overall differences among the groups on all of the cognitive ability and reasoning tasks. Planned comparisons (Tukey's WSD) revealed that, generally, there was no difference between the normative and no change groups and that both of these groups outperformed the nonnormative group. The three groups did not differ in their educational history of mathematics, statistics, and logic courses. Finally, the trends in the data were exactly the same when the analyses were restricted to only those subjects giving the base rate at Step 1.

The pattern of individual differences displayed on the probability assessment tasks was fairly consistent. Individuals who properly adjusted their posterior probabilities in order to take $P(D/{\sim}H)$ into account were more intelligent, more adept at syllogistic reasoning, and more apt to rely on aggregate rather than singular information in statistical reasoning than were individuals whose probability adjustments were in the wrong direction, probably because the latter had ignored—or failed to realize the import of—$P(D/{\sim}H)$. These results converge with those from the information selection task, in which individuals who judged the $P(D/{\sim}H)$ card to be relevant displayed superior cognitive abilities.

Substantial numbers of individuals did not adjust their probabilities at Step 2 in either the David (.25, .70/.90) or the Mark (.80, .40/.05) problem. Cognitively, these subjects resembled the group making the normatively ap-

**Table 1**
**Mean Scores as a Function of Probability Change on Step 2 for Subjects Making Consistent Choices on Both Problems**

| | Group | | | | |
|---|---|---|---|---|---|
| Variable | Nonnormative | No Change | Normative | $F$ Ratio | 95% CI |
| SAT total | 1,069[a] | 1,140[b] | 1,123[b] | 7.41‡ | 14.4–92.7 |
| SAT–M | 551[a] | 586[b] | 592[b] | 6.38‡ | 15.0–68.0 |
| SAT–V | 519[a] | 554[b] | 531 | 3.50* | −13.4–37.5 |
| Vocabulary test | .543[a] | .617[b] | .587 | 3.13* | −.011–.101 |
| Syllogistic reasoning | 17.5[a] | 19.6[b] | 20.3[b] | 8.80‡ | 1.4–4.2 |
| Statistical reasoning | −.493[a] | .498 | .871[b] | 4.14† | .371–2.356 |
| Math background | 1.93 | 2.16 | 2.12 | 1.39 | −.019–.486 |

Note—95% CI = 95% confidence interval for the mean difference between the normative and nonnormative groups. $df = 2{,}169$ for SAT variables, 2,172 for syllogistic reasoning, and 2,174 for the remaining variables; $n$ = 82, 43, and 52 for the nonnormative, no change, and normative groups. $^*p < .05$. $^\dagger p < .025$. $^\ddagger p < .01$. a, b = means with different superscripts are significantly different (Tukey WSD).

propriate adjustment. It is possible that these subjects recognized the relevance of $P$(D/~H), realized that it undercut the apparent diagnosticity of $P$(D/H) taken alone, but were unable to decide how the two balanced each other and thus defaulted to a response of no change. The hypothesis that these subjects did indeed process $P$(D/~H) is bolstered by the fact that they resembled the group who clearly did (the normative group) on several of the other cognitive and reasoning tasks.

Finally, there were some mild indications of convergence in performance on the two Bayesian reasoning tasks. Among individuals who gave the fully normative response on both probability assessment tasks—that is, base rate on Step 1 and the proper directional response to the likelihood ratio on Step 2—69.2% chose $P$(D/~H) in the selection task, whereas this was only true for 45.9% of those who gave one nonnormative response and for 55.5% of those who gave two nonnormative responses [$\chi^2(2) = 6.61, p < .05$].

### Correlational Analyses of Reliance on *P*(H) and *P*(D/~H)

Prior to further exploration of the differential cognitive correlates of the use of base rates and $P$(D/~H), some data reduction was undertaken. The top half of Table 2 displays the results of a principal components analysis conducted on the two sections of the SAT (SAT–M and SAT–V), on the vocabulary test, on the syllogistic reasoning task, and on the statistical reasoning task. Two components had eigenvalues greater than one. The loadings displayed in the table are subsequent to varimax rotation. The first component, which accounted for 41.3% of the variance, appears to be a verbal ability component (high loadings on SAT–V and the vocabulary test). The statistical reasoning task also had

**Table 2**
**Summary of Principal Components Analyses for Experiment 1; Component Loadings for all Variables After Varimax Rotation**

| Variable | Component 1 | Component 2 |
|---|---|---|
| First Analysis | | |
| SAT–M | — | .819 |
| SAT–V | .881 | — |
| Vocabulary test | .852 | — |
| Syllogistic reasoning | — | .762 |
| Statistical reasoning | .416 | .399 |
| % Variance accounted for | 41.3% | 21.8% |

| Variable | Component 1 | Component 2 | Component 3 |
|---|---|---|---|
| Second Analysis | | | |
| SAT–M | — | .802 | — |
| SAT–V | .891 | — | — |
| Vocabulary test | .843 | — | — |
| Syllogistic reasoning | — | .709 | — |
| Statistical reasoning | .429 | .352 | — |
| $P$(D/~H) composite | — | .483 | −.606 |
| Base rate composite | — | — | .858 |
| % Variance accounted for | 31.6% | 16.6% | 15.5% |

Note—Component loadings lower than .250 have been eliminated.

a moderate loading on this component. The second component, which accounted for 21.8% of the variance, might be termed a problem-solving component. It received high loadings from the SAT–M and the syllogistic reasoning task and a moderate loading from the statistical reasoning task, which was the only variable to be factorially complex in that it did not attain simple structure.

Factor scores for the two principal components of this analysis were computed and used to predict performance on two composite scores reflecting sensitivity to $P$(D/~H) and to base rates across the two Bayesian paradigms investigated in Experiment 1. The first composite score, termed the $P$(D/~H) composite, was formed by assigning a score of +1 for choosing the $P$(D/~H) alternative in the information selection task and a score of −1 for not choosing $P$(D/~H). To this score was added the score on Step 2 of the David (.25, .70/.90) problem and the score on Step 2 of the Mark (.80, .40/.05) problem. This step was scored −1 if the subject adjusted his/her probability in a nonnormative direction, 0 if the subject did not adjust his/her probability at all, and +1 if the subject adjusted his/her probability in a normative direction. Thus, scores on the $P$(D/~H) composite index could range from −3 to +3. Scores on the base rate composite variable were formed by simply scoring +1 if $P$(H) was chosen on the selection task, +1 for the base rate response on Step 1 of the David (.25, .70/.90) problem, +1 for the base rate response on Step 1 of the Mark (.80, .40/.05) problem, and 0 otherwise. Thus, scores on the base rate composite index could range from 0 to +3.

A multiple regression analysis was conducted using the $P$(D/~H) composite index as the criterion variable and the verbal ability and problem-solving factor scores as predictor variables. The multiple $R$ (.293) of the regression was statistically significant [$F(2,325) = 15.24, p < .001$], as was each of the standardized beta weights for each of the predictors [verbal ability factor score beta weight = .138, $F(1,325) = 6.73$, $p < .01$; problem-solving factor score beta weight = .258, $F(1,325) = 23.61, p < .001$]. Each factor score was thus an independent predictor, although the problem-solving component score was a somewhat more potent independent predictor. A parallel analysis conducted with the base rate composite index as the criterion variable resulted in a nonsignificant multiple $R$ [.125, $F(2,325) = 2.60$, $.05 < p < .10$], and neither beta weight was a significant independent predictor. It might be argued that there was some restriction of range on base rate usage in the David (.25, .70/.90) and Mark (.80, .40/.05) problems because 84.7% and 81.9% of the subjects, respectively, responded with the base rate on these problems. However, there was no such restriction of range on the information selection task, where 41.5% of the sample chose the base rate and 58.5% did not. A parallel regression analysis run on only base rate usage in the selection task yielded a nonsignificant multiple $R$ [.103, $F(2,335) = 1.81, p > .10$].

The regression analyses thus confirm the results from the discrete, dichotomized analyses presented above, which indicated that the cognitive ability and reasoning measures were significant predictors of sensitivity to $P$(D/~H) but not of base rate usage. Other correlational analyses converge on the same conclusion. For example, the $P$(D/~H) and base rate composite indices were both correlated with a composite index of general cognitive ability formed by standardizing the five cognitive ability and reasoning measures—(SAT–M, SAT–V, the vocabulary test, the syllogistic reasoning task, and the statistical reasoning task—and averaging the standard scores. A test for difference between dependent correlations (see J. Cohen & P. Cohen, 1983, pp. 56–57) revealed that the correlation of the cognitive ability composite with the $P$(D/~H) composite (.259) was significantly higher than that with the base rate composite [.093, $t(337) = 2.08$, $p < .05$].

Finally, the bottom half of Table 2 presents the results of a principal components analysis which included the $P$(D/~H) composite and base rate composite along with the five cognitive and reasoning tasks (a full correlation matrix is presented in the Appendix). Three components had eigenvalues greater than one. The loadings displayed in the table are subsequent to varimax rotation. The structure of the five cognitive and reasoning variables are the same as in the previous analysis displayed in the top half of Table 2. The first component is the verbal ability component, with high loadings on SAT–V and the vocabulary test and a moderate loading on statistical reasoning. The second component is the problem-solving component, with high loadings on SAT–M and the syllogistic reasoning task and a moderate loading from the statistical reasoning task. The $P$(D/~H) composite variable had a moderate loading on this component that was slightly higher than the loading for the statistical reasoning task. The base rate composite variable, in contrast, did not load on either the verbal ability or problem-solving components, but instead formed a third component that was also defined by a negative loading on the $P$(D/~H) composite.

In summary, the results of a variety of different analyses all indicated that the processing of $P$(D/~H) information is more strongly related to cognitive and reasoning ability than is the tendency to process base rate information. This was indicated in analyses focusing on dichotomized choices (Figure 1), in regression analyses, in tests for differences in dependent correlations, and in principal-component analyses. Although it is true that the magnitude of some of the significant effects was small, many of the classifications employed are based on a very few number of trials, and many of the effects may be attenuated by psychometric limitations. For example, the dichotomization of the sample is based on a single administration of the information selection task, yet the effect of picking or not picking the $P$(D/~H) card had an effect size (Cohen's $d$) of .32 on SAT total scores (means of 1,122 and 1,080, respectively). Rosenthal and Rosnow (1991, p. 446) classify such an effect size as midway between "small" and "moderate." It is probably unreasonable to expect larger effect sizes when one variable is defined by a single choice of the subject on a single task. The next experiment will present a direct replication of some of these effects.

Finally, the fact that this experiment was sensitive enough to detect relatively small effects leads us to interpret some of the surprising outcomes with respect to the $P$(H) choice in the selection task as true null findings. For example, the power of the experiment to detect an effect size of .20 was .45, and the power of the experiment to detect an effect size of .30 was .79. In the next experiment, we attempt to replicate these null findings with an even more powerful experiment.

## EXPERIMENT 2

With respect to two of the three critical pieces of information in the selection task, the findings of Experiment 1 were quite divergent as regards patterns of individual differences. Subjects choosing $P$(D/~H) on the selection task tended to have higher cognitive and reasoning abilities than did subjects not making this selection. Sensitivity to $P$(D/~H) on the probability assessment task was also disproportionately a characteristic of the more cognitively able subjects. In contrast, subjects choosing the base rate on the selection task did not display superiority on any other cognitive task. Indeed, there were some mild trends in the opposite direction.

In Experiment 2 we attempt to replicate these intriguing findings regarding the base rate; we also investigate another paradigm in which individuals have been found to ignore or underweight data that are relevant to the nonfocal hypothesis. One difference between the selection task and the probability assessment task is that, in the latter, the subject must actively use $P$(D/~H). Another paradigm that involves even more active processing is the $2 \times 2$ covariation detection paradigm (Wasserman et al., 1990). This is because two components of information (the B and D cells; see the task description below for definitions) must be amalgamated to form $P$(D/~H), and then $P$(D/~H) must be used in a quantitatively normative manner (McKenzie, 1994). Previous research has also indicated that cell D—and hence $P$(D/~H)—is underutilized in this paradigm (Arkes & Harkness, 1983; Kao & Wasserman, 1993; Schustack & Sternberg, 1981; Wasserman et al., 1990). We again address the issue of whether individual differences in this tendency are reliably correlated with other cognitive abilities. In addition, the variant of the covariation task we use allows the assessment of the effects of prior belief on data evaluation (Broniarczyk & Alba, 1994; Klaczynski & Gordon, 1996).

### Method

#### Subjects

The subjects were 611 undergraduate students (186 males and 425 females) recruited through an introductory psychology subject pool at a medium-sized state university. Their mean age was 19.2 years ($SD$ = 2.8). The mean reported verbal SAT score of the 592 students who filled in this part of the demographics questionnaire was 522 ($SD$ = 73), the mean reported mathematical SAT score was 583 ($SD$ = 83), and the mean total SAT score was 1,105 ($SD$ = 127). These reported scores match the averages for this institution (520, 587, and 1,107, respectively) quite closely (Straughn & Straughn, 1995). The same four questions on mathematics/statistics background and the same 0–4 point scale were formulated as in Experiment 1. The mean background score (2.13; $SD$ = .78) was similar to that in Experiment 1 (2.04). No questions on previous logic courses were asked on the demographics questionnaire; however, that variable had no relation to any of the tasks examined in Experiment 1.

### Tasks

**Information selection task.** The information selection task (completed by 596 subjects), syllogistic reasoning task ($n$ = 603), statistical reasoning task ($n$ = 610), and vocabulary test ($n$ = 603) were administered as were those in Experiment 2. The new task was a covariation task which required the assessment of the covariation information in the face of prior beliefs about the relationship.

**Covariation judgment task.** For this task, we adapted a paradigm where people are presented with covariation information that is accommodated by the format of a 2 × 2 contingency table (see Wasserman et al., 1990) and added a belief bias component to it (see Evans, Over, & Manktelow, 1993; Levin et al., 1993). The subjects evaluated 25 contingencies which were embedded within the context of 25 different hypothetical relationships. Each of the 25 problems had two parts. In the first part, subjects were asked their opinions on a hypothetical relationship between two variables. They were then asked to evaluate the degree of association between the two variables in the data of a hypothetical research study. For example, in one problem the subjects were asked whether they believed that couples who live together before marriage tend to have successful marriages. They indicated their degree of agreement with this hypothesized relationship on a scale ranging from −10 (*strongly disagree*) to +10 (*strongly agree*) and centered on 0 (neutral). After responding on this scale, the subjects were told to imagine that a researcher had sampled 250 couples and found that (1) 50 couples did live together and had successful marriages; (2) 50 couples did live together and were divorced; (3) 50 couples did not live together and had successful marriages; and (4) 100 couples did not live together and were divorced.

These data correspond to four cells of the 2 × 2 contingency table traditionally labeled A, B, C, and D (see Levin et al., 1993). Subsequent to the presentation of the data, the subjects were asked to judge the nature and extent of the relationship between living together before marriage and successful marriages in these data on a scale ranging from +10 (*positive association*) to −10 (*negative association*) and centered on 0 (*no association*).

The remaining 24 problems dealt with a variety of hypotheses (e.g., that secondary smoke is associated with lung problems in children; that exercise is associated with a sense of well-being; that eating spicy foods is associated with stomach problems; that being an early-born child is associated with high achievement; that getting chilled is associated with catching a cold; that psychics can help police solve crimes; that watching violent television is associated with violent behavior). The cell combinations used in the 25 problems were based on those listed in Table 2 of Wasserman et al. (1990). The cell values in that table were multiplied by five. The results of each of the hypothetical experiments were listed as above, except that they were always presented in the order cell D, cell C, cell B, cell A, in order to encourage greater reliance on cell D. Previous experiments have indicated that subjects weight the cell information in the order cell A > cell B > cell C > cell D, cell D reliably receiving the least weight and/or attention (see Kao & Wasserman, 1993; Levin et al., 1993; Wasserman et al., 1990). The tendency to ignore cell D is nonnormative, as indeed is any tendency to differentially weight the four cells. The normatively appropriate strategy (see Allan, 1980; Kao & Wasserman, 1993; Shanks, 1995) is to use the conditional probability rule—subtracting the probability of the target hypothesis when the indicator is absent from the probability of the target hypothesis when the indicator is present. Numerically, the rule amounts to calculating the $\Delta p$ statistic: [A/(A+B)] − [C/(C+D)] (see Allan, 1980). For example, the $\Delta p$ value for the problem presented above is +.167, indicating a fairly weak positive association. The $\Delta p$ values used in the 25 problems ranged from −.600 (strong negative association) to .600 (strong positive association). The covariation judgment task was completed by 605 subjects.

### Procedure

The subjects completed the tasks during a single 2-h session in which they completed some other tasks not part of the present investigation. They were tested in small groups of 3 to 8 individuals. Because a few subjects left certain forms blank, fewer than 611 subjects completed some of the tasks.

## Results

### Covariation Judgment Task

Each subject's judgments of the degree of contingency in each of the 25 problems were correlated with the $\Delta p$ value for each problem, with each of the four cell values, and with their agreement with the hypothesis being tested. Twenty-five of the 605 subjects made judgments that failed to correlate significantly with $\Delta p$, with cell A, or with their agreement with the hypotheses. Furthermore, none of these subjects' judgments displayed significant correlations with any of the other individual cells (B, C, or D) or with a linear combination of cells that represented any other well-known strategy (see Kao & Wasserman, 1993). We took this as a sign of either inattention or failure to understand the task and treated the data from these 25 subjects as missing in analyses involving the covariation task.

Of the remaining 580 subjects, only 44 made judgments that were not significantly correlated with $\Delta p$. The mean individual correlation with $\Delta p$ was .639 ($SD$ = .192), and the median correlation was .676. These average individual subject correlations were somewhat lower than those observed by Wasserman et al. (1990) using the same $\Delta p$ values (their median correlation was .816)—probably because our problems involved potential belief bias, whereas theirs did not. Indeed, 299 of our 580 subjects displayed significant correlations between their judgments of contingency and their agreement with the hypothesis being tested. The mean correlation was .363. The latter finding—and the regression analyses presented below—provide evidence that beliefs affect data evaluation, in an analysis where the individual's performance is the unit of analysis. (Levin et al., 1993, observed such an effect in aggregate data.) This paradigm and analysis provide another way of studying belief-bias effects on information processing (C. A. Anderson & Kellam, 1992; Baron, 1991a, 1995; Klaczynski & Gordon, 1996; Koehler, 1993a).

The mean correlations between contingency judgments and the values of cells A, B, C, and D were .568, −.523, −.427, and .401, respectively. The signs of these mean correlations are in the appropriate direction of the normative $\Delta p$ formula: positive for cells A and D and negative for cells B and C. However, the normative strategy also dictates equal weighting of the four cells, and our results converged with previous findings that indicated that subjects typically weight the cells in the order cell A > cell B > cell C > cell D (Kao & Wasserman, 1993; Wasserman et al., 1990). A similar pattern was obtained when each individual's contingency judgments were predicted by a regression equation containing the values of the four cells and the level of the subject's agreement with the hypothesis. Across the 580 analyses, the mean beta weight for cell A was .317, for cell B was −.267, for cell C was −.168, for cell D was .049, and for hypothesis agreement was .234. This analysis likewise indicated proper directional use of the cells but nonnormative unequal weighting. Both the raw correlations and the regression analysis indicated that cell D is underweighted. The regression analysis also indicated the relative magnitude of the belief bias effect in this paradigm. The beta weight for hypothesis agreement was roughly the same as the beta weight for cell B in absolute magnitude, and its independent influence was actually greater than was that of cells C and D.

**Information Selection Task**

The choices observed in the information selection task in Experiment 2 were highly convergent with those observed in Experiment 1. The normatively correct choice of *P*(H), *P*(D/H), and *P*(D/~H) was made by 15.1% of the sample. The most popular choice (43.0% of the sample) was again the two components of the likelihood ratio, *P*(D/H) and *P*(D/~H). The choice of the base rate, *P*(H), and of the numerator of the likelihood ratio, *P*(D/H)—while ignoring the denominator of the likelihood ratio, *P*(D/~H)—was made by 18.5% of the sample, and 19.1% chose *P*(D/H) only. These four patterns accounted for over 95% of the choices in our study. Almost all subjects (96.3%) viewed the *P*(D/H) card as relevant, and very few (2.7%) viewed the *P*(~H) card as relevant. Overall, 63.3% of our subjects chose *P*(D/~H) as a necessary card, and 35.7% of the sample thought it was necessary to know the base rate, *P*(H). However, only 18.3% of the sample viewed *both* the base rate and *P*(D/~H) as relevant.

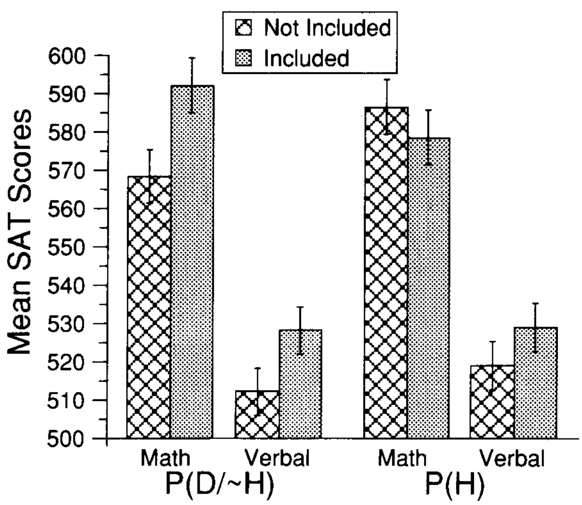


**Figure 2. Mean SAT scores as a function of whether the subjects included or did not include *P*(D/~H) and *P*(H) in their choices on the selection task of Experiment 2. Error bars represent the 95% confidence intervals for the difference between the pairs of sample means. In this figure, the error bars span a mean by a distance equal to the critical value of *t* multiplied by the estimated standard error of the difference between sample means.**

The SAT scores of the subjects who did and of those who did not include *P*(D/~H) in their choices are compared on the left side of Figure 2. Replicating the findings of Experiment 1, the mean SAT–M score of the subjects who selected the *P*(D/~H) card as one of their choices was significantly higher than that of the individuals who did not choose that card [592 vs. 568, $t(576) = 3.32, p < .01$; 95% confidence interval for the difference = 9.7 to 37.9]. The mean SAT–V score was likewise significantly higher [528 vs. 512, $t(576) = 2.53, p < .025$; 95% confidence interval for the difference = 3.6 to 28.4]. Similarly, a converging measure of general ability—the vocabulary test—displayed a significant difference [.511 vs. .479, $t(586) = 2.29, p < .025$; 95% confidence interval for the difference = .005 to .059] favoring those who chose *P*(D/~H). There were also significant differences favoring *P*(D/~H) choosers on the syllogistic reasoning task [17.9 vs. 16.9, $t(586) = 2.84, p < .01$; 95% confidence interval for the difference = 0.32 to 1.75] and on the statistical reasoning task [.275 vs. −.389, $t(593) = 2.91, p < .01$; 95% confidence interval for the difference = .216 to 1.113]. As in Experiment 1, choosing *P*(D/~H) was not associated with educational history. There was no significant difference in mathematics background, as assessed by the mathematics background composite variable.

The left side of Figure 3 presents the comparisons of the two groups on various indices of performance on the covariation judgment task. There were numerous indications of superior processing by the group who chose *P*(D/~H) on the selection task—including several indications that they were more sensitive to cells C and D in their contingency judgments. For example, their judgments displayed significantly higher correlations with the optimal measure of contingency, $\Delta p$ [.658 vs. .612, $t(566) = 2.76, p < .01$; 95% confidence interval for the difference = .013 to .079]. They also displayed higher correlations with the two cells that tend to be underweighted by most subjects in this paradigm—cell C [−.445 vs. −.403, $t(566) = -2.69, p < .01$; 95% confidence interval for the difference = .011 to .072] and cell D [.415 vs. .382, $t(566) = 2.10, p < .05$; 95% confidence interval for the difference = .002 to .062]. Whereas their significantly higher correlation with $\Delta p$ indicated that the subjects in the group choosing *P*(D/~H) on the selection task were more data-driven on the covariation judgment task, their judgments displayed significantly lower correlations with their agreement with the hypoth-

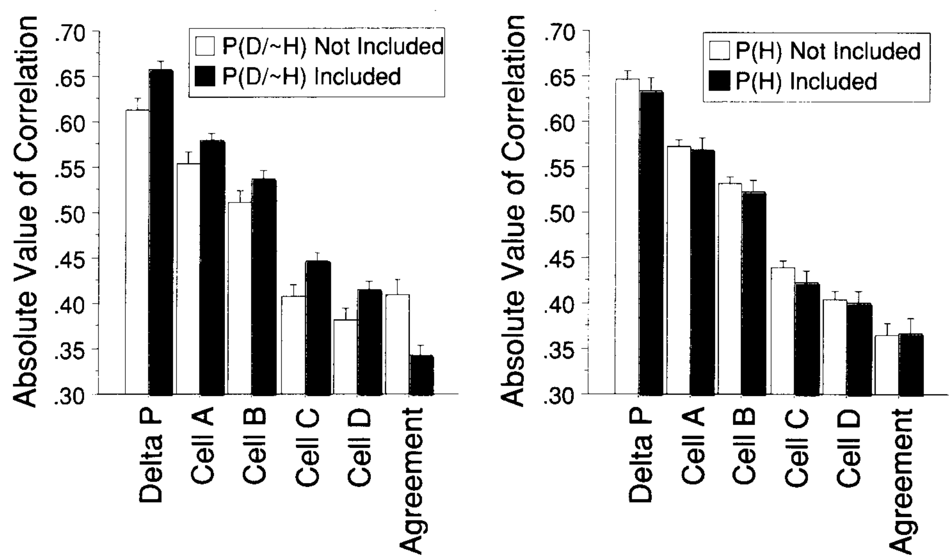


**Figure 3. The mean correlations with components of the covariation task as a function of whether the subjects included or did not include *P*(D/~H) in their choices on the selection task of Experiment 2 are presented on the left. The mean correlations with components of the covariation task of those who included and those who did not include *P*(H) in their choices on the selection task are presented on the right. Error bars represent the 95% confidence intervals for the difference between the pairs of sample means. In this figure, the error bars span a mean by a distance equal to the critical value of *t* multiplied by the estimated standard error of the difference between sample means.**

esis tested [.342 vs. .408, $t(566) = -3.03, p < .01$; 95% confidence interval for the difference = .023 to .110]. This is a further indication that this group was more data-driven in its judgment of contingency. A similar pattern was obtained when the groups were compared on the beta weights obtained when the four cell values and hypothesis agreement were entered as predictors of contingency judgments.

A parallel set of comparisons are presented on the right side of Figures 2 and 3 for those individuals who chose and for those who did not choose the base rate, *P*(H), on the selection task. The results are strikingly different from those obtained when the sample was split on the basis of *P*(D/~H), although they are consistent with the findings of Experiment 1. Specifically, the two groups did not differ significantly on a single variable in Experiment 2. The general cognitive ability (SAT scores, vocabulary test), reasoning abilities, and covariation judgment scores of the subjects who judged the base rate irrelevant in the selection task were no different from the scores of the individuals choosing *P*(H) as one of their cards in the selection task. Ceiling effects do not obscure the interpretation of this outcome, inasmuch as 35.7% of the sample chose the base rate and 64.3% did not. These null findings were not due to a lack of statistical power, because the power of the experiment to detect an effect size as low as .20 was .64, and the power of the experiment to detect an effect size of .30 was .93.

Taken collectively, the results indicate that people who picked *P*(D/~H) were better reasoners on other independent tests but that people who picked the base rate were not. The latter finding held at the extremes of the distribution as well. Of the 59 individuals (roughly 10%) with the lowest SAT scores (mean SAT total of 862), 33.9% chose *P*(H)—exactly the same percentage as that of the 59 individuals with the highest SAT scores (mean SAT total of 1,317).

There was a small but statistically significant negative association between the choice of *P*(H) and of *P*(D/~H); 69.9% of the subjects who did not choose *P*(H) did choose *P*(D/~H), whereas only 51.1% of the subjects who chose *P*(H) also chose *P*(D/~H) [$\chi^2(1) = 20.01, p < .001$]. As in Experiment 1, there were differences favoring the subjects who chose *P*(D/~H) but not *P*(H) over those who chose *P*(H) but not *P*(D/~H) on all the tasks. Finally, the 90 individuals who made the normatively appropriate choice—[*P*(H), *P*(D/H), and *P*(D/~H)]—had significantly higher SAT–V and vocabulary scores than the rest of the sample and achieved higher scores on the other cognitive variables, but the latter differences did not attain statistical significance.

Individual differences in covariation judgment performance were examined by partitioning the sample on the basis of a median split of their weighting of cell D. As Table 3 indicates, subjects giving relatively higher weighting to cell D had significantly higher SAT scores, vocabulary scores, syllogistic reasoning scores, and statistical reasoning scores than did subjects giving relatively lower weighting to cell D. These differences were not due to differences in mathematics background. Thus, the results from the covariation judgment task converged with those from the information selection task in indicating that sensitivity to *P*(D/~H) was associated with greater facility on other cognitive and reasoning tasks.

**Table 3**
**Mean Scores (With Standard Deviations) of Subjects Giving a High Weighting to Cell D in the Covariation Judgment Task ($n$ = 300) and Those Giving a Low Weighting ($n$ = 279)**

| Variable | Low *M* | Low *SD* | High *M* | High *SD* | *t* value | 95% CI |
|---|---|---|---|---|---|---|
| SAT total | 1,084 | 131 | 1,133 | 117 | 4.68* | 28.3–69.3 |
| SAT–M | 570 | 87 | 600 | 76 | 4.47* | 17.2–44.2 |
| SAT–V | 515 | 71 | 533 | 73 | 2.98* | 6.2–30.0 |
| Vocabulary test | .484 | .170 | .521 | .153 | 2.73* | .010–.063 |
| Syllogistic reasoning | 16.7 | 4.1 | 18.4 | 4.3 | 4.88* | 1.0–2.4 |
| Statistical reasoning | −.359 | 2.6 | .340 | 2.7 | 3.15* | .264–1.135 |
| Mathematics background | 2.10 | .76 | 2.16 | .76 | 0.89 | −.068–.180 |

Note—95% CI = 95% confidence interval for the mean difference. $df$ = 562 for SAT variables, 571 for the vocabulary test and syllogistic reasoning, and 577 for statistical reasoning and math background. *$p < .01$, two-tailed.

### Correlational Analyses of Reliance on $P$(H) and $P$(D/~H)

The top half of Table 4 displays the results of a principal components analysis conducted on the two sections of the SAT (SAT–M and SAT–V), the vocabulary test, the syllogistic reasoning task, and the statistical reasoning task (a full correlation matrix is presented in the Appendix). The loadings displayed in the table are subsequent to varimax rotation. The first component is the problem-solving component (high loadings on SAT–M, syllogistic reasoning, and statistical reasoning) and the second is the verbal ability component (high loadings on SAT–V and the vocabulary test). The structure of the five cognitive and reasoning variables is the same as that in the previous analysis conducted on the Experiment 1 data (displayed in the top half of Table 2), except that the components are reversed. One small difference was that the statistical reasoning task loaded solely on the problem-solving component in Experiment 2 but displayed moderate loadings on both components in Experiment 1.

Factor scores for the two principal components of this analysis were computed and used to predict performance on a composite score reflecting sensitivity to $P$(D/~H). The composite score was formed by assigning a score of 1 for choosing the $P$(D/~H) alternative in the information selection task and a score of 0 for not choosing $P$(D/~H). To this score was added a score of 1 if the subject displayed a significant correlation with cell D on the covariation judgment task and a score of 0 if this correlation was not significant. Thus, scores on the $P$(D/~H) composite index could range from 0 to +2.

A multiple regression analysis was conducted using the $P$(D/~H) composite index as the criterion variable and the verbal ability and problem-solving factor scores as predictor variables. The multiple $R$ (.243) of the regression was statistically significant [$F(2,538) = 16.83, p < .001$], as was each of the standardized beta weights for each of the predictors [verbal ability factor score beta weight = .100, $F(1,538) = 5.77, p < .025$; problem solving factor score beta weight = .221, $F(1,538) = 27.81$, $p < .001$]. Each factor score was thus an independent predictor, although, as in Experiment 1, the problem-solving component score was a somewhat more potent independent predictor. A parallel analysis conducted with the choice of the base rate on the information selection task as the criterion variable resulted in a nonsignificant multiple $R$ [.087, $F(2,538) = 2.04, p > .10$], and neither beta weight was a significant independent predictor.

The regression analyses thus confirm the results from the discrete, dichotomized analyses present above, which indicated that the cognitive ability and reasoning measures were significant predictors of sensitivity to $P$(D/~H) but not of base rate usage. Other analyses converge on the same conclusion. For example, a composite index of general cognitive ability was formed by standardizing the five cognitive ability and reasoning measures and averaging the standard scores. A test for difference between dependent correlations (J. Cohen & P. Cohen, 1983, pp. 56–57) revealed that the correlation of the ability composite with the $P$(D/~H) composite (.238) was significantly higher than that with the base rate choice in the selection task [.022, $t(565) = 3.58, p < .001$].

**Table 4**
**Summary of Principal Components Analyses for Experiment 2; Component Loadings for All Variables After Varimax Rotation**

| Variable | Component 1 | Component 2 |
|---|---|---|
| *First Analysis* | | |
| SAT–M | .736 | — |
| SAT–V | — | .817 |
| Vocabulary test | — | .903 |
| Syllogistic reasoning | .786 | — |
| Statistical reasoning | .602 | — |
| % Variance accounted for | 44.9% | 18.6% |
| *Second Analysis* | | |
| SAT–M | .642 | — |
| SAT–V | .747 | .282 |
| Vocabulary test | .652 | .320 |
| Syllogistic reasoning | .657 | — |
| Statistical reasoning | .537 | — |
| Correlation: Cell D | .373 | — |
| $P$(D/~H), selection task | .271 | −.642 |
| $P$(H), selection task | — | .736 |
| % Variance accounted for | 29.1% | 15.1% |

Note—Component loadings lower than .250 have been eliminated.

Finally, the bottom half of Table 4 presents the results of a principal components analysis which included the cell D correlation in the covariation task and $P(D/{\sim}H)$ and $P(H)$ from the selection task, along with the five cognitive and reasoning tasks. The loadings displayed in the table are subsequent to varimax rotation of the first two principal components, both of which have eigenvalues greater than one. Here, the first principal component is largely a general factor with high loadings (>.500) on all five cognitive ability and reasoning tasks. Attention to cell D in the covariation task and the choice of $P(D/{\sim}H)$ in the selection task have positive loadings on this general ability component. However, the choice of the base rate on the selection task has no loading on the general reasoning component but instead serves to largely define the second component [along with a negative loading from $P(D/{\sim}H)$]. This outcome largely converges with the results of the principal components analysis of Experiment 1, presented at the bottom of Table 2.

## DISCUSSION

Across these two experiments, a consistent pattern of individual differences was associated with the choice of $P(D/{\sim}H)$ on two different tasks. Individuals including the $P(D/{\sim}H)$ card in their set of choices on the selection task had significantly higher SAT scores, as well as higher scores on a converging measure of cognitive ability (the vocabulary test). They scored higher on a syllogistic reasoning task and tended to weight more heavily the more reliable aggregate information when making an inductive inference. They performed more optimally on the covariation judgment task of Experiment 2, where they tended to be more data-driven. That is, they were significantly more reliant on the $\Delta p$ of the data and significantly less reliant on their agreement with the hypothesis being tested.

Convergent patterns were observed on the probability assessment task in Experiment 1, where the respondent had to actively use $P(D/{\sim}H)$ to arrive at a correct Bayesian probability adjustment. Here again, individuals who adjusted their probabilities in the right direction on both problems—and thus must have processed $P(D/{\sim}H)$—outperformed the individuals making a misadjustment on both problems—probably because of a failure to process $P(D/{\sim}H)$—on all of the cognitive ability and reasoning measures. Finally, subjects giving relatively higher weighting to cell D in the covariation judgment task in Experiment 2 demonstrated more efficient processing on the cognitive ability and reasoning tasks.

Both Experiments 1 and 2 revealed a markedly divergent pattern when individuals were classified according to whether they included the base rate, $P(H)$, in their set of choices in the selection task. People who included the base rate in their set of choices did not differ in cognitive or reasoning ability from individuals who deemed the base rate irrelevant, nor did they differ in their degree of data-driven processing on the covariation judgment task.

Many factors undoubtedly affect the choice of both $P(D/{\sim}H)$ and the base rate. For example, the use of the latter has been related to various pragmatic factors (see Koehler, 1996), such as the origin of the individuating evidence with which it is combined (see, e.g., Schwarz et al., 1991), whether it is presented first or last (Krosnick, Li, & Lehman, 1990), and whether the base rate is varied in a within-subjects design (Fischhoff, Slovic, & Lichtenstein, 1979). Similarly, the tendency to process $P(D/{\sim}H)$ is affected by whether subjects must evaluate it in conjunction with a $P(D/H)$ value or singularly (Doherty et al., 1996), whether subjects are evaluating an asymmetric (present/absent) variable (Beyth-Marom, 1982), and whether subjects have experienced a nondiagnostic likelihood ratio (Doherty et al., 1996). The influence of these factors suggests that subjects may be primarily making relevance judgments in the task rather than reasoning probabilistically (Sperber et al., 1995). Thus, it is likely that choosing either component is affected by a variety of pragmatic reasoning schemas and conversational implicatures (Adler, 1984; Cheng & Holyoak, 1985; Hilton, 1995; Levinson, 1995; Macchi, 1995; Sperber et al., 1995). Few of these factors may reflect analytic (see Evans, 1984, 1989, 1996) probabilistic reasoning of the type that might be linked to computational capacity. The influence of these factors probably accounts in part for the small effect size for choosing $P(D/{\sim}H)$. However, in the case of $P(H)$, there was no evidence at all of the operation of a probabilistic reasoning process linked to computational capacity.

Numerous correlational analyses confirmed the pattern of a greater covariance with other cognitive and reasoning tasks for $P(D/{\sim}H)$ than for $P(H)$. In short, $P(D/{\sim}H)$ enters into a pattern of a positive manifold with other cognitive tasks, whereas $P(H)$ does not. These patterns of individual differences are interestingly convergent with patterns found in the history of disputes surrounding the use of base rates as opposed to the likelihood ratio. Despite occasional warnings that an exclusive focus on falsification is not always efficacious (see, e.g., Friedrich, 1993; Klayman & Ha, 1987, 1989; see also J. R. Anderson, 1990, p. 159), virtually all theorists agree that both components of the likelihood ratio are critical for optimal probability adjustment. In contrast, many issues surrounding the use of base rates continue to provoke controversy and critical comment. These issues range from arguments that inappropriate normative models have been assumed (Birnbaum, 1983; L. J. Cohen, 1979, 1981, 1982, 1986; Gigerenzer, 1991, 1993; Gigerenzer et al., 1991; Kyburg, 1983; Levi, 1983, 1996) to arguments that many problems used in psychological research are linguistically and pragmatically unclear to subjects (Gigerenzer et al., 1988; Gigerenzer & Hoffrage, 1995; Hilton, 1995; Koehler, 1996; Macchi, 1995; Macdonald, 1986; Margolis, 1987; Schwarz et al., 1991).

Our results mirror the differential controversy surrounding these two components of probabilistic reasoning. With respect to the component that is least controversial, *P*(D/~H), we found that people who tend to use it also reason efficiently and properly in other domains. In contrast, the most efficient and cognitively able reasoners in the study were *not* more likely to choose the base rate in the selection task—precisely the component that has been the subject of the most theoretical dispute. Given that a positive manifold is expected among most cognitive tasks (Carroll, 1993), the latter finding might be interpreted as indicating that the wrong normative model is being applied to the task (Birnbaum, 1983; Gigerenzer, 1993; Stein, 1996) or that there are alternative construals of the task that are equally likely to be chosen by the more able or less able subjects. We prefer the latter alternative, in part because, if the former were the case, we might—given the arguments about normative models outlined in the introduction—have expected a significant correlation in the direction counter to the base rate. That is, a positive manifold with *whatever* alternative response was normative might be expected. Because this did not happen, it is perhaps more likely that there are alternative construals of the task that are equally rational (Berkeley & Humphreys, 1982; Broome, 1990; Hastie & Rasinski, 1988; Hilton, 1995; Politzer & Noveck, 1991). For example, Zukier and Pepitone (1984) found that being told to approach a task as a clinician decreased reliance on base rates. The use of the word *diagnose* in the selection-task instructions might have helped to trigger such a case-specific reasoning schema.

In summary, when the present results are interpreted within an adaptionist framework, they lead to the conclusion that the tendency to evaluate *P*(D/~H) is part of the underlying probabilistic model that people use to optimize their tracking of the world. However, our analysis of individual differences suggests that the status of the base rate in rational models appears to be different from that of *P*(D/~H). As mentioned in the introduction, critics of the normative models that are assumed in many psychological experiments have charged that there is "no support for the view that people would choose in accord with normative prescriptions if they were provided with increased capacity" (Lopes & Oden, 1991, p. 209). There now *is* such evidence with respect to *P*(D/~H). However, this criticism appears to be somewhat on the mark with respect to judgments of base rate relevance in the selection paradigm employed in these experiments.

## REFERENCES


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

ALLAN, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, **15**, 147-149.

ANDERSON, C. A., & KELLAM, K. (1992). Belief perseverance, biased assimilation, and covariation detection: The effects of hypothetical social theories and new data. *Personality & Social Psychology Bulletin*, **18**, 555-565.

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

ANDERSON, J. R. (1991). Is human cognition adaptive? *Behavioral & Brain Sciences*, **14**, 471-517.

ANDERSON, R. C., & FREEBODY, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (Ed.), *Advances in reading/language research* (Vol. 2, pp. 231-256). Greenwich, CT: JAI Press.

ARKES, H. R., & HARKNESS, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, **112**, 117-135.

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 (Ed.), *Insights into decision making: A tribute to Hillel J. Einhorn* (pp. 200-216). Chicago: University of Chicago Press.

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

BARON, J. (1991a). Beliefs about thinking. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 169-186). Hillsdale, NJ: Erlbaum.

BARON, J. (1991b). Some thinking is irrational. *Behavioral & Brain Sciences*, **14**, 486-487.

BARON, J. (1995). Myside bias in thinking about abortion. *Thinking & Reasoning*, **1**, 221-235.

BELL, D., RAIFFA, H., & TVERSKY, A. (EDS.) (1988). *Decision making: Descriptive, normative, and prescriptive interactions*. Cambridge: Cambridge University Press.

BERKELEY, D., & HUMPHREYS, P. (1982). Structuring decision problems and the "bias heuristic." *Acta Psychologica*, **50**, 201-252.

BEYTH-MAROM, R. (1982). Perception of correlation reexamined. *Memory & Cognition*, **10**, 511-519.

BEYTH-MAROM, R., & FISCHHOFF, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality & 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.

BLESS, H., STRACK, F., & SCHWARZ, N. (1993). The information functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology*, **23**, 149-165.

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.

BRATMAN, M. E., ISRAEL, D. J., & POLLACK, M. E. (1991). Plans and resource-bounded practical reasoning. In J. Cummins & J. Pollock (Eds.), *Philosophy and AI: Essays at the interface* (pp. 7-22). Cambridge, MA: MIT Press.

BRONIARCZYK, S., & ALBA, J. W. (1994). Theory versus data in prediction and correlation tasks. *Organizational Behavior & Human Decision Processes*, **57**, 117-139.

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.

CAMPBELL, D. T. (1987). Evolutionary epistemology. In G. Radnitzky & W. W. Bartley (Eds.), *Evolutionary epistemology, rationality, and the sociology of knowledge* (pp. 47-89). La Salle, IL: Open Court.

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

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

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

CHENG, P. W., & HOLYOAK, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, **17**, 391-416.

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

COHEN, J., & COHEN, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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 & 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. (1986). *The dialogue of reason*. Oxford: Oxford University Press.

Cooksey, R. W., & Freebody, P. (1987). Aspects of a computer-managed test of children's reading vocabulary: Reliability, validity, and characterization of knowledge. *Reading Psychology*, **8**, 103-118.

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

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

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

Deary, I. J. (1995). Auditory inspection time and intelligence: What is the direction of causation? *Developmental Psychology*, **31**, 237-250.

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

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

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

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). Pseudodiagnosticity. *Acta Psychologica*, **43**, 111-121.

Eells, E., & Maruszewski, T. (Eds.) (1991). *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science*. Amsterdam: Editions Rodopi.

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

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, **32**, 53-88.

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. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, **87**, 223-240.

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

Evans, J. St. B. T., Over, D. E., & Manktelow, K. (1993). Reasoning, decision making and rationality. *Cognition*, **49**, 165-187.

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

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior & Human Performance*, **23**, 339-359.

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

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, **100**, 298-319.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence. *Psychological Science*, **7**, 237-241.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, **98**, 254-267.

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 & 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.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

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

Halberstadt, N., & Kareev, Y. (1995). Transitions between modes of inquiry in a rule discovery task. *Quarterly Journal of Experimental Psychology*, **48A**, 280-295.

Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology*, **101**, 252-254.

Hastie, R., & Rasinski, K. A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 193-208). Cambridge: Cambridge University Press.

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

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

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

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

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

Kao, S. F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1363-1386.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan.

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

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 211-228.

Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 596-604.

Koehler, J. J. (1993a). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior & Human Decision Processes*, **56**, 28-55.

Koehler, J. J. (1993b). The normative status of base rates at trial. In N. J. Castellan (Ed.), *Individual and group decision making* (pp. 137-149). Hillsdale, NJ: Erlbaum.

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

Koehler, J. J., & Shaviro, D. N. (1990). Veridical verdicts: Increasing verdict accuracy through the use of overtly probabilistic evidence and methods. *Cornell Law Review*, **75**, 247-275.

Kornblith, H. (Ed.) (1985). *Naturalizing epistemology*. Cambridge, MA: MIT University Press.

Kornblith, H. (1993). *Inductive inference and its natural ground*. Cambridge, MA: MIT University Press.

Krosnick, J. A., Li, F., & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality & Social Psychology*, **59**, 1140-1152.

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

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 & Human Decision Processes*, **56**, 331-347.

LEVI, I. (1983). Who commits the base rate fallacy? *Behavioral & Brain Sciences*, **6**, 502-506.

LEVI, I. (1996). Fallacy and controversy about base rates. *Behavioral & Brain Sciences*, **19**, 31-32.

LEVIN, I. P., WASSERMAN, E. A., & KAO, S. F. (1993). Multiple methods of examining biased information use in contingency judgments. *Organizational Behavior & Human Decision Processes*, **55**, 228-250.

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

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.

MACDONALD, R. (1986). Credible conceptions and implausible probabilities. *British Journal of Mathematical & 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.

MARKOVITS, H., & NANTEL, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, **17**, 11-17.

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

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

MCKENZIE, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, **26**, 209-239.

NEWELL, A. (1982). The knowledge level. *Artificial Intelligence*, **18**, 87-127.

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

NEWSTEAD, S., POLLARD, P., EVANS, J. S., & ALLEN, J. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, **45**, 257-284.

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

NISBETT, R. E. (1993). *Rules for reasoning*. Hillsdale, NJ: Erlbaum.

NISBETT, R. E., & ROSS, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

OAKSFORD, M., & CHATER, N. (1993). Reasoning theories and bounded rationality. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 31-60). London: Routledge.

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

OAKSFORD, M., & CHATER, N. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking & Reasoning*, **1**, 121-152.

POLITZER, G., & NOVECK, I. A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research*, **20**, 83-103.

POULTON, E. C. (1994). *Behavioral decision theory: A new approach*. Cambridge: Cambridge University Press.

PYLYSHYN, Z. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.

ROSENTHAL, R., & ROSNOW, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

SCHOEMAKER, P. (1991). The quest for optimality: A positive heuristic of science? *Behavioral & Brain Sciences*, **14**, 205-245.

SCHUM, D. (1990). Discussion. In R. M. Hogarth (Ed.), *Insights into decision making: A tribute to Hillel J. Einhorn* (pp. 217-223). Chicago: University of Chicago Press.

SCHUSTACK, M. W., & STERNBERG, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, **110**, 101-120.

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.

SHAFER, G. (1988). Savage revisited. In D. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions* (pp. 193-234). Cambridge: Cambridge University Press.

SHAFIR, E. (1991). On the applicability of a rational analysis to human cognition. *Behavioral & Brain Sciences*, **14**, 502-503.

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

SHANKS, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology*, **48A**, 257-279.

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

SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.

SPERBER, D., CARA, F., & GIROTTO, V. (1995). Relevance theory explains the selection task. *Cognition*, **57**, 31-95.

STANKOV, L., & DUNN, S. (1993). Physical substrata of mental energy: Brain capacity and efficiency of cerebral metabolism. *Learning & Individual Differences*, **5**, 241-257.

STANOVICH, K. E. (in press). *Variable rationality*. Mahwah, NJ: Erlbaum.

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

STICH, S. (1990). *The fragmentation of reason*. Cambridge, MA: MIT Press.

STRAUGHN, C. S., & STRAUGHN, B. L. (EDS.) (1995). *Lovejoy's college guide* (23rd ed.). New York: Simon & Schuster.

THAGARD, P. (1982). From the descriptive to the normative in philosophy and logic. *Philosophy of Science*, **49**, 24-42.

THAGARD, P., & NISBETT, R. E. (1983). Rationality and charity. *Philosophy of Science*, **50**, 250-267.

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.

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

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

VERNON, P. A., & MORI, M. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence*, **16**, 273-288.

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, & Cognition*, **16**, 509-521.

WEST, R. F., & STANOVICH, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science*, **2**, 325-330.

WHITE, T. G., SLATER, W. H., & GRAVES, M. F. (1989). Yes/no method of vocabulary assessment: Valid for whom and useful for what? In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction* (38th Yearbook of the National Reading Conference, pp. 391-397). Chicago: National Reading Conference.

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

ZIMMERMAN, J., BRODER, P. K., SHAUGHNESSY, J. J., & UNDERWOOD, B. J. (1977). A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition. *Intelligence*, **1**, 5-31.

ZUKIER, H., & PEPITONE, A. (1984). Social roles and strategies in prediction: Some determinants in the use of base-rate information. *Journal of Personality & Social Psychology*, **47**, 349-360.

## APPENDIX

**Intercorrelations Among the Primary Variables in Experiment 1**

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. SAT–V | | | | | | |
| 2. SAT–M | .12 | | | | | |
| 3. Vocabulary test | .61 | .13 | | | | |
| 4. Syllogistic reasoning | .24 | .34 | .25 | | | |
| 5. Statistical reasoning | .31 | .13 | .21 | .27 | | |
| 6. $P$(D/~H) composite | .14 | .24 | .13 | .18 | .17 | |
| 7. Base rate composite | .07 | .08 | .07 | .10 | .02 | −.14 |

Note—Correlations larger than .11 are significant at the .05 level (two-tailed).

**Intercorrelations Among the Primary Variables in Experiment 2**

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. SAT–V | | | | | | | |
| 2. SAT–M | .32 | | | | | | |
| 3. Vocabulary test | .56 | .24 | | | | | |
| 4. Syllogistic reasoning | .35 | .40 | .24 | | | | |
| 5. Statistical reasoning | .27 | .23 | .20 | .28 | | | |
| 6. Correlation: Cell D | .14 | .23 | .13 | .21 | .08 | | |
| 7. $P$(D/~H), selection task | .11 | .14 | .09 | .12 | .12 | .09 | |
| 8. $P$(H), selection task | .07 | −.05 | .05 | −.01 | −.01 | −.01 | −.19 |

Note—Correlations larger than .09 are significant at the .05 level (two-tailed).

(Manuscript received July 17, 1996;
revision accepted for publication January 6, 1997.)