Information Seeking on Bayesian Conditional Probability Problems: A Fuzzy-trace Theory Account

CHRISTOPHER R. WOLFE
Miami University, USA

ABSTRACT
Recently, the 'heuristics and biases' approach to the study of decision making has been criticized, with a call for better integrated theory. Three experiments stemming from fuzzy-trace theory addressed information seeking on probability problems, and the cognitive representation of hit-rates, base-rates, and the contrapositive. As predicted by the fuzzy-trace principle of 'denominator neglect', many subjects exhibited 'conversion errors', confusing the hit-rate, P(A|B), with the answer, P(B|A). These subjects sought base-rates less often than other subjects. On causal problems, more subjects correctly represented base-rates, sought base-rates more often, and produced more accurate estimates than on non-causal problems. Subjects tutored on the meaning of the hit-rate sought the base-rate more often, and were more accurate than control subjects. Results are explained by fuzzy-trace theory principles of gist extraction, fuzzy processing preference, denominator neglect, and output interference.

KEY WORDS information seeking; Bayesian; base-rate; fuzzy-trace

Recently, theory and empirical findings arising from the 'heuristics and biases' approach to the study of decision making have been called into question, particularly with respect to the use of base-rate information (Funder, 1994; Koehler, 1993, 1994; Macchi, 1994). For example, Koehler (1993, p. 3.2) claims that there has been 'a vast oversale of the so-called "base-rate fallacy" in the probabilistic judgment literature'. Indeed, Koehler argues that the base-rate fallacy is a 'myth' and that the heuristics and biases paradigm, exemplified by Kahneman and Tversky's (1972) representative heuristic, has not been supported by subsequent data. Koehler (1993, p. 4.2) argues that more recent 'data were ignored and the theory persisted [because] the underlying principle was too attractive to abandon on account of data'. Along with these criticisms has been a call for more ecologically valid research, and better integrated theoretical perspectives (Koehler, 1993, 1994; Koonce, 1993).

Phrases such as 'base-rate fallacy', 'ignore base-rates', and 'base-rate neglect' have been criticized because they do not accurately describe psychological processes. It is unclear, for example, whether people underweight or ignore base-rate data. As Hamm (1994, p. 2) notes, the term base-rate neglect erroneously 'implies a flaw in an attentional process, so that insufficient weight is given to base-rate information'. Key issues that have not been adequately addressed in the literature concern the kinds of information people seek in making probability judgments, and the nature of the cognitive representation of probabilistic information such as base-rates and the contrapositive.

CCC 0894-3257/95/020085-24
© 1995 by John Wiley & Sons, Ltd.

Received 17 December 1993
Accepted 27 October 1994
The research reported here assesses the cognitive representation of probabilistic information, and distinguishes between a tendency to underweigh base-rates, and a failure to seek base-rate information at all. This research stems from a new theory of human cognition with broad ramifications for the study of decision making and reasoning, and strong empirical support: fuzzy-trace theory (Brainerd and Reyna, 1990a,b; Reyna, 1991; Reyna and Brainerd, 1990, 1991, 1992, in press a,b; Reyna et al., 1990).

A key feature of fuzzy-trace theory is gist representation. The basic claim is that when information is encoded, global gist-like patterns, impressions, and essences are encoded along with verbatim information. The result is a multifaceted fuzzy-to-verbatim representation of information. Individual knowledge items are represented along a continuum such that vague, fuzzy-traces co-exist with more precise verbatim representations. Moreover, people exhibit a strong preference to reason with the vaguest gist-like representations allowable for a given task.

Fuzzy-trace theory has broad implications of human judgment and decision making. The first implication is that subjects prefer to act on gist-like representations of problems rather than faithful representations of the givens — even when verbatim responses are memorially accessible (Reyna and Brainerd, 1991). Reyna et al. (1990, p. 7) propose that ‘the natural habit of thought is to manipulate the barest sense of ideas, a fluid and uncertain process, rather than to march rigidly from given facts to inescapable conclusions’. Thus, just ‘because subjects can discriminate differing quantities, and can act on those discriminations, it does not follow that problems are invariably solved by processing information at the highest possible resolution’ (Reyna et al., 1990, p. 13). Indeed fuzzy-trace theory predicts that subjects will prefer to reason on the basis of representations at the fuzzy end of the continuum. These processes are generally adaptive and even rational within boundaries (Simon, 1956; Reyna et al., 1990), yet they produce reliable and predictable deviations from rational models in some situations.

The preference for fuzzy representations and fuzzy processing proposed by fuzzy-trace theory is distinct from the ‘constructiveness preference’ and ‘cognitive economy’ suggested by information-processing models such as the contingent process perspective (Payne, 1982; Payne et al., 1992). Cognitive economy maintains that people act as if a ‘calculation’ is made about the cost of processing information and the need for accuracy. From this perspective, task performance is the result of ‘the interaction between the limited memory and computational capabilities of decision makers and the complexity of task environments’ (March, 1978, quoted in Payne et al., 1992). Fuzzy-trace theory suggests cognitive processes that generally act upon fuzzy problem representations, and specifically argues against memorial and computational limits as central explanatory constructs. Fuzzy-trace theory holds that the ‘degree of fuzziness’ changes from one situation to the next, but the preference for fuzziness is pervasive. A fuzzy processing preference is, perhaps, ‘economical’ from an evolutionary perspective, in that many routine tasks require only a fuzzy representation. Moreover, reasoning with fuzzy representations sometimes improves performance over reasoning with verbatim information. Cognitive economy, from an information-processing perspective, also leads to predictions, such as memory dependency, that have been disconfirmed by fuzzy-trace research (Brainerd and Reyna, 1989, 1990a).

Phenomena explained by memory dependency in other models are generally explained by the fuzzy-trace principle of output interference. This principle states that the act of generating overt (or internal) responses interferes with reasoning by creating system-wide noise that degrades performance (Brainerd and Reyna, 1989, 1990a; Reyna and Brainerd, 1989). Output interference is analogous to the difficulty of learning to juggle. Juggling is difficult because the act of throwing one ball interferes with the act of catching another. One learns to juggle not by adding more hands (resources) but by learning to coordinate catching and throwing (outputs). Output interference suggests significant debilitating consequences for processing irrelevant information, or even too much relevant information. Fuzzy-trace theory predicts that subjects will seek relatively little information, even if that information is readily available and the cost of information access is negligible.

Assessing subjects’ gist representations may provide insights into performance on a wide variety of
logical and statistical reasoning tasks. According to fuzzy-trace theory, failure to extract the relevant gist (misperceiving the task situation) is a major source of errors in quantitative problem solving (Reyna and Brainerd, in press b). The purpose of the current set of experiments was to determine how people represent the kinds of information required to solve Bayesian conditional probability problems, and the relationship between those representations and information seeking behavior. Bayesian problems are of broad interest because, in modified forms, the kinds of information used in Bayes’ Theorem are necessary for many types of cognitive and social psychological processes including attribution (Lipe, 1991), logical deduction (Cummins et al., 1991, Revlis, 1975), induction (Fratianne and Cheng, 1991; Freedman, 1992; Mynatt et al., 1977; Platt and Griggs, 1992), and probabilistic inference (Edwards, 1968; Kahneman and Tversky, 1973, 1982; Klar, 1991).

Bayes’ Theorem states that

\[
P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(\neg B) \cdot P(A|\neg B)}
\]

For example, the Ball and Jar problem in Appendix 1 may be represented in Bayes’ Theorem as P(Red Ball | Black Jar) =

\[
P(\text{Black Jar} | \text{Red Ball}) \cdot P(\text{Red Ball})
\]

\[
\frac{P(\text{Black Jar} | \text{Red Ball}) \cdot P(\text{Red Ball}) + (\text{Not Red Ball}) \cdot P(\text{Black Jar} | \text{White Ball})}{P(\text{Black Jar} | \text{Red Ball}) \cdot P(\text{Red Ball}) + (\text{Not Red Ball}) \cdot P(\text{Black Jar} | \text{White Ball})}
\]

The paradigm employed in this research asked subjects to estimate a conditional probability P(B|A) and gave them a menu of available information including P(A|B), P(B), and P(A|\neg B). In this case, the P(Black Jar | Red Ball) is available to subjects as item 1 on the information menu: ‘Of all the red balls drawn in this game, the percentage that came from the Black Jar is 30%.’ The P(Red Ball) is available as item 5 on the information menu as ‘The percentage of balls drawn from all of the jars in this game that are red is 70%’, P(Not Red Ball) is readily inferred from this. Similarly, the P(Black Jar | White Ball) is available as item 3 in the information menu as ‘Of all the white balls drawn in this game, the percentage that came from the black jar is 50%’. The solution, approximately 58%, may be computed as

\[
0.5833 = \frac{0.3 \cdot 0.7}{0.3 \cdot 0.7 + 0.3 \cdot 0.5}
\]

Fuzzy-trace theory predicts that subjects will prefer to reason with simplified representations of information. A powerful, yet normatively incorrect, simplification ‘strategy’ is conversion (Reyna and Brainerd, 1991), in which subjects convert conditional propositions to biconditional propositions (Hamm, 1993, 1994; Koehler, 1994; Macchi, 1994). There is a good deal of evidence that people sometimes interpret statements of the form ‘all A are B’ to mean ‘all B are A’ and ‘P implies Q’ to mean ‘Q implies P’, (Chapman and Chapman, 1959; Revlis, 1975; Revlin and Leirer, 1980; Wilkins, 1928). Pollatsek et al. (1987) found that subjects also confuse P(A|B) with P(A and B). Fuzzy-trace theory predicts that a large number of subjects will produce conversion errors in their reasoning about the hit-rate and the contrapositive on conditional probability problems, i.e. confusing the hit-rate P(A|B) with P(B|A), and the contrapositive P(A|\neg B) with P(\neg B|A).

According to fuzzy-trace theory, conversion errors are the result of fuzzy representations formed at
the time of encoding, along with more precise representations. Thus, All As are Bs' may be represented at the fuzzy end of the continuum as something like 'As = Bs' or 'As and Bs go together' (see Klar, 1990, 1991, for a compatible alternative account). Given the nature of output interference, subjects reasoning with this level of gist are likely to ignore other kinds of information — including useful information such as base-rates and the contrapositive. Thus, as Hamm (1993) demonstrated the 'base-rate fallacy' may stem, in part, from subjects' tendency to confuse P(A|B) with P(B|A).

From the perspective of fuzzy-trace theory, conditional probability judgments are difficult because it is confusing to work with multiple, nested inclusion relationships. Reyna (1991) and Reyna and Brainerd (in press a, b) explain difficulties in reasoning on a number of tasks involving part–whole relations as the product of 'denominator neglect'. Performance on tasks as diverse as Piagetian class-inclusion problems, conjunction problems, and conditional probability problems may be explained by subjects' tendency to 'ignore' normatively relevant denominators. As Reyna and Brainerd (in press b, p. 28) note, people 'often base probability judgments on comparisons between numerators, and, thereby, avoid having to keep track of the relationship between those numerators and the denominators in which they are included'.

Exhibit 1. 2 × 2 contingency table for the Balls and Jars problem representing the source and color of balls

<table>
<thead>
<tr>
<th></th>
<th>Red ball drawn</th>
<th>White ball drawn</th>
<th>Totals for jars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drawn from black jar</td>
<td>Cell I</td>
<td>Cell II</td>
<td>Cell I + II</td>
</tr>
<tr>
<td>Drawn from other jars</td>
<td>Cell III</td>
<td>Cell IV</td>
<td>Cells III + IV</td>
</tr>
<tr>
<td>Totals for balls</td>
<td>Cells I + III</td>
<td>Cells II + IV</td>
<td>Cells I + II + III + IV</td>
</tr>
</tbody>
</table>

To clarify the denominator neglect associated with conditional probability judgments, it is useful to examine the contingency table provided as Exhibit 1. This illustrates the contingency relationships addressed by Bayes' Theorem, using the example of the Ball and Jar problem in Appendix 1. In this problem, there are red and white balls drawn from black or other jars. The goal is to estimate the probability of drawing a red ball from a black jar, P(B|A). In the contingency table, this is cell I divided by cells I plus II. Subjects may choose to examine the hit rate, P(A|B), which is the probability that the jar is black given that a red ball is drawn. In the contingency table, this is cell I divided by cells I plus III. Another piece of available information is the contrapositive, P(¬A|¬B), which is the probability that the jar is black given that a white ball is drawn. In the contingency table, this is cell II divided by cells II plus IV. The only other piece of relevant information available to subjects is the base rate, P(B), which is the overall probability of drawing a red ball from any jar. In the contingency table, this is cells I plus III divided by cells I plus II plus III plus IV.

The tendency to ignore denominators has several consequences for Bayesian tasks. First, subjects are likely to confuse P(A|B) with P(B|A) when the denominators (cells I + III and cells I + II) are ignored, since both share the numerator cell I. Thus, subjects will be most interested in information about cell I, which in the case of the Ball and Jar problem, is information about instances of red balls being drawn from black jars. However, to reduce the complexities associated with output interference, subjects will exclude the normatively important denominators from their representations, resulting in conversion errors in their reasoning. Second, even subjects who understanding the meaning of the base rate are likely to consider it less central than the hit rate due to denominator neglect. Subjects may erroneously reason that P(B) is less important than P(A|B) because the latter is specifically about instances of red balls being drawn from black jars, while the former is about red balls being drawn from any jar. Finally, denominator neglect suggests that subjects will confuse the contrapositive, P(¬A|¬B), with P(¬B|A) since the contrapositive is cell II divided by cells II plus IV, and P(¬B|A) is cell II divided by cells I plus II.
Although the use of base-rates has been studied extensively (Ajzen, 1977; Bar-Hillel, 1980; Calogero and Nelson, 1992; Doherty et al., 1979; Kahneman and Tversky, 1982; Klayman and Ha, 1987; Lynch and Ofir, 1989; Medin and Bettger, 1991), central questions have not been addressed. One of the most fundamental questions concerns the way people represent base-rate information. Currently, little is known about peoples’ qualitative representation of statistical concepts. A second issue concerns the locus and extent of base-rate neglect. One version of the base-rate fallacy is that, while making probability estimates, many people fail to consider base-rates. A weaker claim made by Lynch and Ofir (1989, p. 170) is that the “base-rate fallacy” is observed only when (a) one combines base and case cues that lead to dissimilar judgments when each is considered alone, and (b) the case cue is high in numeric value.

Fuzzy-trace theory holds that people maintain multiple representations of information. Thus determining which representation was ‘in play’ during reasoning is no simple matter. The technique employed in the current work to assess subjects’ gist representations was to ask them to describe the meaning of each piece of information and its relevance to the problem after problem solving. Asking subjects to describe stimuli in their own words is a common approach to assessing gist in research on memory and comprehension (e.g. Adams et al., 1990; Langer et al., 1987; Torgesen et al., 1988). However, in this research, memorial limitations were not an issue since each item was presented verbatim at the time of gist assessment. This technique also has the advantage of producing tangible data which may be categorized by independent scorers, and compared to overt behavior ‘blind’ to task performance.

A great deal of research has focused on factors affecting the use of base-rate and contrapositive information. Ajzen (1977) found that subjects were more likely to employ base-rates when they were causally related to the criterion. Bar-Hillel (1980) found that base-rates are perceived as more relevant when causally linked to the judged outcome. There is also a large body of evidence suggesting that people frequently ignore contrapositive information (e.g. McGill and Klein, 1993). Mynatt et al. (1977, 1978) found that subjects demonstrated a strong tendency to seek evidence confirming their hypotheses, and seldom sought information that would falsify their hypotheses. Mynatt et al. described this as a ‘confirmation bias’. Moreover, Beyth-Marom and Fischhoff (1983) found that most subjects who sought the contrapositive did so for the wrong reasons. Klayman and Ha (1987) reconceptualized the ‘confirmation bias’ in terms of a positive test strategy or heuristic. The positive test strategy (Fratianne and Cheng, 1991; Freedman, 1992) is the tendency to test cases that are expected (or known) to have the property of interest rather than those expected (or known) to lack that property (Klayman and Ha, 1987, p. 211). Previous research has not determined how a causal relationship affects the way people represent Bayesian information, specifically whether causal relationships affect judgments by affecting the representation of base-rates, hit-rates, or the contrapositive.

The central issues investigated in this research are (1) what kinds of information people seek in making conditional probability judgments, and (2) how people represent base-rate, hit-rate, and contrapositive information. By studying information seeking, we are able to address fundamental questions that cannot be addressed using problems for which all of the information is provided (see Beyth-Marom and Fischhoff, 1983; Fischhoff and Beyth-Marom, 1983). For example, although previous research indicates that subjects ‘ignore’ base-rate information (Bar-Hillel, 1980), it is not clear whether ‘ignore’ means that subjects underweigh such information or whether they fail to consider it at all.

**EXPERIMENT 1**

It was predicted that subjects would seek out hit-rate information significantly more often than either the base-rate or contrapositive. Verbal descriptions of the relevance of each piece of information were collected from each subject, and categorized by the experimenter. Special attention was given to the
extent to which subjects expressed a correct representation of the base-rate, claimed that base-rates are irrelevant, and provided evidence of conversion errors, confusing P(A|B) with P(B|A) and P(A|¬B) with P(¬B|A).

A secondary issue concerned the impact of irrelevant information on information seeking and utilization. Previous research demonstrates that irrelevant information affects information seeking (Johnson, 1988), and the use of base-rates (Manis et al., 1980; Tversky and Kahneman, 1982). However, the ability to ignore irrelevant information is also a hallmark of cognitive development (Reyna and Brainerd, in press a). Irrelevant items were created by taking the first clause in sentences describing relevant items, and changing the second clause. These items also provide a control for a simple matching heuristic (i.e. seeking sentences containing the phrase ‘red balls’).

Method

Subjects
Forty-eight undergraduate introductory psychology students served as subjects on a voluntary basis, and fulfilled a course requirement by participating. Few (if any) had formal training in statistics.

Design and materials
Each subject received six conditional probability problems that were versions of the Ball and Jar problem (Appendix 1). Each problem included a statement of the problem and a menu of available information, including statements describing the base-rate, P(B), hit-rate, P(A|B), and contrapositive, P(A|¬B). Subjects were randomly assigned to one of two experimental conditions. One half were presented with only relevant information in the menu, and the rest were also presented with three irrelevant pieces of information.

After seeing the numeric values associated with a piece of information (e.g. P(A|B): 'Of all the red balls drawn in this game, the percentage that came from the black jar is 30%'), subjects were instructed to make an estimate of the conditional probability P(B|A) before selecting another piece of information or going on to the next problem. Subjects were permitted to seek out as much or as little of the information in the menu as they desired. They were free to examine information in any order of their choosing. This design yielded two groups of 24 subjects.

The six problems represent the six possible orders of the base-rate, hit-rate, and contrapositive on the menu. The numeric values 30%, 50%, and 70% were always assigned to either the base-rate, hit-rate, or contrapositive. The six problems also represent the six possible orders of these three numeric values. Other than these differences, the six problems differed only in the colors of the balls and jars. As a counterbalancing measure, values and problems were matched with each other using two 6 × 6 Latin squares to insure that each subject received all six orders of P(B), P(A|B), and P(A|¬B) as well as all six orders of the values 30%, 50%, and 70%. This counterbalanced the effects of problem order, value set, and menu order in a completely balanced incomplete blocks design. The two 6 × 6 Latin squares yielded twelve sets of six problems, insuring that every subject received each value set and each menu order only once. Subjects were randomly assigned to one of these sets within each condition.

Procedure
Subjects were run individually on an IBM AT computer. On a practice problem, the experimenter read the problem aloud, and ‘walked’ subjects through their options. Subjects worked at their own pace. All
the information, including the subjects' estimates, remained on the screen as long as they worked on a problem. Next, they were given the gist-assessment task. The task asked them to describe the meaning of each piece of information in the ball and jar problem and to describe its relevance to the problem.

Exhibit 2. Information seeking: mean frequency of selecting the base-rate P(B), hit-rate P(A|B), contrapositive P(A|¬B) and irrelevant information P(I) over six problems (N = 48)

| Condition                     | P(B) | P(A|B) | P(A|¬B) | P(I) |
|-------------------------------|------|--------|---------|------|
| Relevant information only     | 3.46 | 5.00   | 3.50    |      |
| Irrelevant and relevant info  | 3.83 | 4.58   | 3.13    | 0.62 |
| Total                         | 3.65 | 4.79   | 3.32    | 0.62 |

Results

Exhibit 2 presents the mean frequency of selection for base-rate, hit-rate, contrapositive, and irrelevant information. A Dunn’s Critical Difference test indicates that there are no significant differences between the relevant and irrelevant information conditions, CD = 0.96. When given a choice, subjects chose to examine any of the three irrelevant pieces of information on only 10% of the problems. Over the six problems, subjects chose to examine the hit-rate significantly more often than either the base-rate, or the contrapositive, CD = 0.96, p < 0.01. Twenty-seven per cent (27%) of the subjects examined only one piece of information, 43% two pieces of information, and only 30% all three pieces of relevant information.

Gist descriptions

Verbal data from the gist description task were categorized into a small number of discrete categories. The experimenter was blind to condition and to performance on the task (indeed, categorization was done before any of the quantitative analyses). A second rater independently categorized a randomly selected sample of verbal data for 32 subjects after receiving training on data from 16 subjects. Responses from the 16 subjects used for training produced prototypes of categorization criteria. Inter-rater reliability was 0.82 and disagreements were settled by consensus. There were no significant differences between experimental groups in their verbal descriptions.

A large number of subjects (54%) had an appropriate representation of the base-rate, and only 5% said the base-rate was irrelevant. However, only a small fraction of the subjects were able to adequately describe the true relevance of the hit-rate (14%), and an even smaller number (6%) correctly described the relevance of the contrapositive. A large number of subjects (34%) exhibited evidence of a conversion error (Eddy, 1982; Reyna and Brainerd, 1991), confusing P(A|B) with P(B|A). Similarly, a number of subjects (41%) provided evidence of the conversion error P(A|¬B) = P(¬B|A).

Information seeking and gist representation

The last set of analyses examined information seeking as a function of subjects’ gist representations. Subjects making the P(A|B) = P(B|A) conversion error sought out base-rate information significantly less often over six problems, compared to all other subjects, 2.79 times (46%) vs. 4.00 (67%), t(46) = 1.91,
A similar analysis examined the frequency of searching for the base-rate as a function of verbal descriptions of the contrapositive. Subjects making the P(A|¬B) = P(¬B|A) conversion error sought out base-rate information significantly less often over six problems, compared to all other subjects, 3.00 times (50%) vs. 4.00 (67%), t (46) = 1.71, p < 0.05. No other differences in information seeking were significantly predicted by verbal responses. These findings suggest that base-rate neglect stems, in part, from conversion errors. Because half of the subjects making the conversion error with the hit-rate also exhibit the conversion error for the contrapositive, it was not possible to fully distinguish these two effects in the first experiment.

Discussion
A majority of subjects (54%) indicated an appropriate gist representation of the meaning of the base-rate. For example, one subject said ‘Gives you an idea of what chances of pulling out a red ball no matter what color jar you have’. Another subject said ‘Of all the balls drawn, the chances of it being red’. A third said ‘The relative number of red balls to white balls overall’. Only 5% said that this information was irrelevant. This finding is at odds with Bar-Hillel’s (1980) claim that subjects ignore base-rates because they believe them to be irrelevant, and consistent with Hamm’s (1993) results. Although the subjects who said the base-rate was irrelevant examined the base-rate less than other subjects, those expressing a correct representation of the base-rate sought base-rate information on only 62% of the problems. Thus, even subjects who understand the meaning of the base-rate often ignore base-rates.

Many subjects (34%) made a conversion error in their reasoning about the hit-rate. For example, one subject said ‘This is the chance I have for drawing a red ball from the black jar’. Another subject said ‘This tells you your chances of winning, exactly’. Another said ‘This is what the problem above is asking for’. Several subjects (41%) also showed evidence of a conversion error in their verbal descriptions of the contrapositive. For example, one subject said ‘This is my chance of picking a white ball from the black jar’. Another said ‘This shows you the chances of losing’.

Even when presented with a menu of only three pieces of information, subjects rarely examined all the available information. The presence of irrelevant information did not lead to a reduction in the seeking of relevant information. Indeed, the small nonsignificant trend was in the opposite direction. Subjects examined any of the three pieces of irrelevant information only 10% of the time.

A major finding from the first experiment was that conversion errors are associated with the base-rate fallacy. Consistent with fuzzy-trace theory, subjects who exhibited a conversion error in either their verbal descriptions of the hit-rate or the contrapositive examined the base-rate significantly less often than all other subjects.

EXPERIMENT 2

The second experiment examined the effects of causal relationships and prior expectations on Bayesian conditional probability problems. Some problems described situations in which there is a causal relationship between the base-rate and the target case, while others did not. Second, some problems described situations for which people have a general idea about the magnitude of the base-rates, while others did not.

Fuzzy-trace theory suggests that the process of gist extraction often involves reducing causally relevant information to gist. In many contexts, developing useful gist necessitates representing causally relevant information. Thus, people have a great deal of experience extracting causally relevant information and incorporating it into their gist representations. On causal problems, it is predicted that subjects will generate appropriate verbal gist representations of the base-rates without specific reference.
to probability theory or relative frequency. For example, the gist of the base-rate on the tennis problem in Appendix 2 ('Of all the matches that Jean has ever played, the percentage that Jean has won is 90%') might be expressed in English as 'Jean is a good tennis player'. It is predicted that it will be more difficult for subjects to develop appropriate gist representations of the base-rate on non-causal problems. When problems describe situations in which there is a causal relationship between the base rate and the target case, subjects will be more likely to understand and seek base-rate information. When subjects incorporate base-rates into their estimates, they are less likely to rely exclusively on the hit-rate, and more likely to be accurate. Thus, it was predicted that when the base-rate was causally related to the target, it would be selected more often than when it was not.

As McKenzie (1994) notes, there has been confusion in the literature about the relationship between people's prior beliefs and base-rates provided by the experimenter. These differences were addressed in Experiment 2 by comparing behavior on problems for which subjects had prior expectations about the base-rates with those for which they did not. Prior expectations about base-rates were expected to make base-rate information more familiar and thus more likely to be selected.

It was predicted that subjects demonstrating an appropriate gist representation of the base-rate would seek out base-rate information significantly more often than other subjects. A second prediction was that conversion errors would be associated with a lower frequency of seeking base-rate information. Because subjects have considerable experience reducing causally relevant information to gist representations, it was predicted that significantly more subjects would demonstrate a correct representation of the base-rate on problems describing a causal relationship.

These data were also analyzed with respect to the overall accuracy of the subjects relative to Bayes' Theorem. Because causally relevant information is typically preserved in gist representations, it was predicted that subjects would be significantly more accurate when the base-rate was causally related to the target case. Subjects were also expected to be more accurate on problems for which they had prior expectations about the base-rates.

Method

Subjects
Ninety-six undergraduate students from the same subject pool used in Experiment 1 served as subjects on a voluntary basis.

Design and materials
Each subject received four problems, each of which included a problem statement and a menu of Bayesian information. The four problems were designed to manipulate perceptions of causation and prior expectations about the base-rate, in a $2 \times 2$ design. Of the four problems, two described situations in which there is a causal relationship while the other two did not. Crossed with this factor, two of the problems described situations for which people have prior expectations about the base-rate, while the other two did not.

In developing these materials, perceptions of causation and prior expectations were measured in a pilot study. One hundred subjects answered questions about causation and prior expectations for several problems. Four problems were selected as best-fitting criteria for perceived causation and prior expectations (i.e. high--high, high--low, low--high, and low--low). The Ball and Jar problem (Appendix 1) rated low on causation and prior expectations. The Tennis Match problem (Appendix 2) rated high on causation and low on prior expectations. The Dormitory problem (Appendix 3) rated low on causation and high on prior expectations. The Test problem (Appendix 4) rated high on causation and prior expectations.
Four sets of numeric values were matched with \( P(B), P(A\|B), P(A\|\neg B), \) and \( P(B|A). \) The four numeric value sets, the four problems, and four problem orders were manipulated in a \( 4 \times 4 \) Greco-Latin Square. This was done to counterbalance any effects of problem order and to ensure that each value set was associated with each problem 24 times. In addition, for each problem, three orders of the base-rate, the hit-rate, and the contrapositive were presented to control for menu order effects. Finally, one-half of the subjects received only relevant information while the other half of the subjects were also presented with irrelevant information. Subjects were randomly assigned to condition and run in four blocks of 24 subjects.

**Procedure**
The procedure was identical to that of Experiment 1.

**Results**
The general strategy for analyzing the data from Experiment 2 was first to test for any effects of counterbalancing measures. When no reliable differences were found, these variables were ignored in subsequent analyses. For all measures, there were no reliable effects for Block, Group, Menu order, or numeric values (this was expected since these values were initially hidden from subjects). Thus, subsequent analyses of information-seeking behaviour collapsed across the range of numeric values. The presence of irrelevant information in the menu also had no effects on processing. There were only two (2) instances of 576 possible where subjects chose to examine irrelevant information. In both of these cases, the subjects did not change their estimates of \( P(B|A) \) after examining irrelevant information or differ appreciably in their use of relevant information. For this reason, the presence of irrelevant information was also ignored in subsequent analyses. This yields a \( 2 \times 2 \) within-subject design comparing the presence or absence of a causal relationship between the base-rate and the target case, and the presence or absence of prior expectations about the base-rate. Although each problem may have idiosyncratic characteristics, only results that were robust across both causal and prior expectation problems are main effects.

**Information seeking**
The first analysis concerns the seeking of base-rate information. Exhibit 3 presents the frequency of subjects seeking base-rates for each of the four problems. A Cochran's \( Q \)-test indicates that subjects sought base-rate information significantly more often when the base-rate was causally related to the target case than when it was not causally related, \( Q = 4.5, p < 0.034 \). There were no reliable effects for prior expectations, \( Q < 1 \), or idiosyncratic effects for individual problems.

---

1 The numeric values (associated with \( P(B|A) = 58\%, 73\%, 82\%, \) and \( 91\% \)), were selected to meet a variety of criteria. In each case, the given values were multiples of 10%. Each value was at least ten percentage points distant from each of the other values in the same set, at least ten percentage points from the values associated with the same kind of information in other value sets, and at least ten points away from the correct value of \( P(B|A) \). The numeric values of the hit-rate were \( 20\%, 30\%, 40\%, \) and \( 70\% \), providing an additional test of Lynch and Ofir's (1989) claim that people ignore base-rates only when the case cue is assigned a high numeric value. This wide range and 'spread' of values make the interpretation of the results clearer. It also reduces any 'carryover' effects from one problem to the next, and promotes more reliable generalizations across a wide range of values. The correct values of \( P(B|A) \) were approximately ten points different from each other, and as close as possible to multiples of 10%. This was done to control for the 'whole number bias' in estimates of the \( P(B|A) \) (Lopes, 1987, p. 171). To ensure that values of \( P(B) \) and \( P(B|A) \) were roughly consistent with prior expectations for the Dormitory and Test problems, these values ranged between \( 60\% \) and \( 90\% \).
Exhibit 3. Percentage of subjects seeking the base-rate P(B) by problem (N = 96)

<table>
<thead>
<tr>
<th>Causally related</th>
<th>Expectations</th>
<th>No expectations</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>59%</td>
<td>Tennis</td>
<td>60%</td>
</tr>
<tr>
<td>Dorm</td>
<td>42%</td>
<td>Ball</td>
<td>46%</td>
</tr>
<tr>
<td>Total</td>
<td>51%</td>
<td>56%</td>
<td>53%</td>
</tr>
</tbody>
</table>

Of the 96 subjects, 81% examined the hit-rate on the Test problem, 92% on the Tennis problem, 89% on the Dormitory problem, and 91% on the Ball problem. There were no significant main effects for causal relationship, $Q < 1$, or for prior expectations, $Q < 1$. However, subjects examined the hit-rate significantly less often on the Test problem than on any of the others, $Q = 6.25, p < 0.01$.

Concerning the frequency of seeking contrapositive information of the 96 subjects, 70% examined the contrapositive on the Test problem, 67% on the Tennis problem, 41% on the Dormitory problem, and 58% on the Ball problem. Subjects were more likely to seek out the contrapositive when there was a causal relationship between the base-rate and the target case, $Q = 8.33, p < 0.004$. There were no effects for prior expectations, $Q < 1$. However, subjects examined the contrapositive significantly less often on the Dormitory problem than on the Ball problem, $Q = 7.41, p < 0.007$. Overall, subjects examined the hit-rate more often (88%) than either the base-rate (53%) or the contrapositive (59%).

**Probability estimates**

The accuracy of subjects’ probability estimates may be expressed in terms of the fit between subjects’ final estimates of $P(B|A)$ and those provided by Bayes’ Theorem. Accuracy is measured here as both the mean difference between the subjects’ final estimates of $P(B|A)$ and Bayesian estimates (S-B) and as the mean absolute value of the difference between subjects’ final estimates and Bayesian estimates (|S-B|). Because subjects were more accurate with some value sets than with others, separate analyses were conducted for each of the four value sets (see Exhibit 4). This was done by dividing the Greco-Latin square by value set, yielding four $2 \times 2$ between-subject ANOVAs measuring the effects of causal relations and prior expectations on accuracy for both S-B ans. (S-B). The ANOVAs are orthogonal to one another in that subjects appear in different conditions in each of the four analyses.

For S-B, subjects were significantly more accurate on causal problems for each of the four numeric value sets. Subjects were significantly more accurate on prior expectation problems only with value set 1 and value set 4. There were no significant interactions. For |S-B|, subjects were significantly more accurate on causal problems with value set 2 and value set 3. Subjects were only significantly more accurate on prior expectation problems with value set 4. Only with value set 1 was there an interaction effect. Subjects were significantly more accurate on the test problem (causal, prior expectations). These data indicate that subjects were generally more accurate on causal problems and sometimes more accurate on prior expectation problems.

Exhibit 4 also presents the mean, median, and mode of the final estimates of $P(B|A)$ produced by subjects for each problem and value set. An examination of the median responses confirms that a causal relationship between the base-rate and the target case is associated with more accurate estimates. More importantly, subjects generally used the hit-rate as their estimates of $P(A|B)$ for non-causal problems, but less often for causal problems. It is also apparent that subjects deviated sharply from Bayes’ Theorem across the full range of problems and value sets. The grand mean of |S-B| was 27.5%. 
Exhibit 4. Average final estimates of \( P(B|A) \) by problem and numeric value set and accuracy relative to Bayes' Theorem \(^a\) \( (n = 24 \text{ per average}) \)

<table>
<thead>
<tr>
<th>Value set 1</th>
<th>Value set 2</th>
<th>Value set 3</th>
<th>Value set 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(A</td>
<td>B) = 30% )</td>
<td>( P(A</td>
<td>B) = 40% )</td>
</tr>
<tr>
<td>( P(B) = 70% )</td>
<td>( P(B) = 80% )</td>
<td>( P(B) = 90% )</td>
<td>( P(B) = 60% )</td>
</tr>
<tr>
<td>( P(A</td>
<td>\neg B) = 50% )</td>
<td>( P(A</td>
<td>\neg B) = 60% )</td>
</tr>
<tr>
<td>( P(B</td>
<td>A) = 58% )</td>
<td>( P(B</td>
<td>A) = 73% )</td>
</tr>
</tbody>
</table>

| Problem | Test (causal, expectations) | | |
|---------|---------------------|------|
| Mean    | 54                  | 52   | 63  | 79 |
| Median  | 50                  | 50   | 77.5| 90 |
| Mode    | 50                  | 75   | 20  | 90 |
| S-B     | -4.1                | -18.9| -19.8| -15.1|
| [S-B]   | 14.2                | 22.3 | 28.3| 16.9|

| Problem | Tennis (causal, no expectations) | | |
|---------|---------------------------------|------|
| Mean    | 38                              | 47  | 49 | 69 |
| Median  | 37.5                            | 45  | 52 | 70 |
| Mode    | 50                              | 40  | 20 | 70 |
| S-B     | -21.5                           | -26.3| -33.5| -22.2|
| [S-B]   | 24.3                            | 27.0 | 35.2| 22.6|

| Problem | Ball (not causal, no expectations) | | |
|---------|-----------------------------------|------|
| Mean    | 37                               | 43  | 36 | 60 |
| Median  | 30                               | 40  | 27.5| 62.5|
| Mode    | 20                               | 40  | 20 | 70 |
| S-B     | -21.1                            | -30.5| -46.4| -30.7|
| [S-B]   | 21.6                             | 30.5 | 46.4| 30.7|

| Problem | Dorm (not causal, expectations) | | |
|---------|---------------------------------|------|
| Mean    | 41                              | 41  | 38 | 73 |
| Median  | 30                              | 40  | 27.5| 70 |
| Mode    | 30                              | 40  | 20 | 70 |
| S-B     | -18.5                           | -31.1| -47.3| -17.6|
| [S-B]   | 22.9                            | 31.8 | 47.8| 17.6|

\(^a\)The accuracy of subjects' final estimates is expressed as the mean of subject's estimate of \( P(B|A) \) minus the Bayesian estimate of \( P(B|A) \) (S-B) and the mean of the absolute value of subject's estimate of \( P(B|A) \) minus the Bayesian estimate of \( P(B|A) \) (S-B). Probabilities are expressed in percentages.

**Gist descriptions of \( P(B) \)**

After completing the information-seeking experiment, subjects were given the gist description task. For 96 subjects describing the 12 pieces of irrelevant information there were only nine instances of subjects describing them as anything but irrelevant. No further analyses were conducted for verbal descriptions of any of the irrelevant information. A second rater independently categorized the relevant verbal data for a randomly selected one-third of the subjects \( (n = 32) \) after receiving training on a randomly selected one-third of the data. Responses from the one-third of the data used for training provided prototypes of categorization criteria. Overall inter-rater reliability was 0.82 for all judgments, and disagreements were settled by consensus.

With respect to gist descriptions of the base-rate, an important distinction is whether subjects expressed an appropriate representation of the concept. Responses were categorized as appropriate if they indicated an overall or unconditional probability. With regards to this distinction, the inter-rater reliability was 0.91. Exhibit 5 indicates the number of subjects correctly describing the meaning of the base-rate for each of the four problems. Cochran's \( Q \)-test indicates that significantly more subjects
Exhibit 5. Percentage of subjects producing correct verbal descriptions of \( P(B) \) by problem \((N = 96)\)

<table>
<thead>
<tr>
<th></th>
<th>Expectations</th>
<th>No expectations</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causally related</td>
<td>Test 63%</td>
<td>Tennis 73%</td>
<td>68%</td>
</tr>
<tr>
<td>Not causally related</td>
<td>Dorm 38%</td>
<td>Ball 36%</td>
<td>37%</td>
</tr>
<tr>
<td>Total</td>
<td>50%</td>
<td>55%</td>
<td>52%</td>
</tr>
</tbody>
</table>

correctly described the meaning of the base-rate on problems describing a causal relationship, \( Q = 45.47, p < 0.0001 \). There were no significant effects for prior expectations or evidence of an interaction.

**Gist descriptions of \( P(A|B) \)**

Only a small number of subjects (14%) indicated an appropriate representation of the hit-rate. There were no significant differences between problems for the frequency of correct verbal responses, \( Q_s < 1 \). A large number of subjects showed evidence of a conversion error on each problem, confusing \( P(A|B) \) with \( P(B|A) \). With regards to this distinction, inter-rater reliability was 0.83. There were no significant differences among the problems for frequency of conversion error, \( Q_s < 1 \). Seventy-four subjects (77%) made a conversion error on at least one problem, and thirty-eight (40%) made the conversion error on three or four problems. This indicates that the error was pervasive, and not unique to only a few subjects or a single problem.

**Gist descriptions of \( P(A|\neg B) \)**

Only a small number of subjects indicated an appropriate representation of \( P(A|\neg B) \). There were no significant differences between problems for the frequency of correct verbal responses, \( Q_s < 1 \). A large number of subjects confused \( P(A|\neg B) \) with \( P(\neg B|A) \) on each problem. With regard to this distinction, the inter-rater reliability was 0.84. Subjects did not make significantly more conversion errors when the base-rate was causally related to the target case, \( Q = 3.46, p < 0.063 \), or for problems for which they had prior expectations about the base-rates, \( Q = 2.31, p < 0.128 \). However, subjects made significantly more conversion errors on the Test problem than on the Dormitory problem, \( Q = 10.26, p < 0.01 \).

A large number of subjects showed evidence of conversion errors in reasoning about the hit-rate and the contrapositive. Both kinds of conversion error tend to be made by the same subjects. The degree of association between the two errors is expressed in four chi squares, comparing the presence or absence of each kind of conversion error, one for each problem. For the Tennis problem, \( \chi^2(1) = 23.50, p < 0.0001 \), for the Ball problem, \( \chi^2(1) = 35.61, p < 0.0001 \), for the Test problem, \( \chi^2(1) = 32.96, p < 0.0001 \), and for the Dormitory problem, \( \chi^2(1) = 13.55, p < 0.0002 \).

**Information seeking and gist representation**

These analyses examined the frequency of selecting the base-rate as a function of whether subjects demonstrated an appropriate representation of the base-rate in their verbal descriptions of the Tennis, Ball, Test, and Dormitory problems. Subjects generally demonstrating a correct representation of the base-rate examined the base-rate more often than those who did not: on the Tennis problem, \( \chi^2(1) = 3.53, p < 0.06 \); on the Ball problem, \( \chi^2(1) = 6.77, p < 0.009 \); on the Test problem, \( \chi^2(1) = 10.02, p < 0.002 \); and on the Dormitory problem, \( \chi^2(1) = 11.70, p < 0.0006 \). Thus, subjects who exhibited an accurate representation of the base-rate tended to seek out base-rate information.
The next four analyses examined the frequency of selecting the base-rate as a function of whether subjects provided evidence of making the conversion error in their gist descriptions of the hit-rate on the Tennis, Ball, Test, and Dormitory problems. Subjects who made the conversion error in their verbal descriptions of the hit-rate examined the base-rate relatively less often than those who did not: on the Tennis problem, $\chi^2(1) = 4.50, p < 0.034$; on the Ball problem, $\chi^2(1) = 2.64, p < 0.10$; on the Test problem, $\chi^2(1) = 5.67, p < 0.017$; and on the Dormitory problem, $\chi^2(1) = 4.91, p < 0.027$. Taken collectively, these results indicate that subjects who confused the hit-rate $P(A|B)$ with $P(B|A)$ sought out the base-rate significantly less often than other subjects.

The relationship between gist descriptions of the contrapositive and information seeking was examined in the next set of analyses. No significant relationships between conversion errors and information seeking were found for the Ball, Test and Dormitory problems, $\chi^2$s < 1. The Tennis problem yielded a marginally significant effect, $\chi^2(1) = 3.80, p < 0.051$. This result is particularly interesting in light of the frequent coincidence of the two types of conversion error.

Conversion error versus understanding $P(B)$

Because both a correct representation of the base-rate and a hit-rate conversion error affect the seeking of base-rate information, efforts were made to distinguish between these two influences. One approach is to determine the extent to which the two types of verbal response were made by the same subjects, by looking for nonsignificant chi squares to verify the independence of the two effects. The conversion error and correct representation of the base-rate were generally not related: on the Tennis problem, $\chi^2(1) = 3.26, p < 0.071$; on the Ball problem, $\chi^2(1) < 1$; on the Test problem, $\chi^2(1) < 1$; and on the Dormitory problem, $\chi^2(1) = 5.42, p < 0.02$. These results indicate that the two types of verbal response were generally independent of one another. Another approach was to examine the influence of a correct representation of the base-rate separately for subjects who made the conversion error and for those who did not. A symmetrical analysis examined the influence of the conversion error on seeking base-rates for subjects who had a correct representation of the base-rate separately from those subjects who did not exhibit a correct representation of the base-rate. These analyses completely control for the influence of the other variable. However, they also greatly reduce statistical power by reducing the number of subjects in each analysis. This approach yielded sixteen separate analyses. Of the 16 $\chi^2$ analyses, eight were significant at $p < 0.05$. The data for all four problems are presented in Exhibit 6.

These results indicate that an appropriate representation of the base-rate is generally associated with an increase in seeking base-rate information. A second independent effect is that the $P(A|B) = P(B|A)$ conversion error is generally associated with a reduction in seeking base-rate information.

Exhibit 6. Percentage of subjects selecting $P(B)$ by conversion error $P(A|B) = P(B|A)$ and correct verbal description of $P(B)$ over four problems (ratio selecting $P(B)$ in parentheses)

<table>
<thead>
<tr>
<th>Description of P(B)</th>
<th>No conversion error</th>
<th>Conversion error</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>76% (82/108)</td>
<td>59% (55/93)</td>
<td>68% (137/201)</td>
</tr>
<tr>
<td>Not correct</td>
<td>49% (48/98)</td>
<td>24% (20/85)</td>
<td>37% (68/183)</td>
</tr>
<tr>
<td>Total</td>
<td>63% (30/206)</td>
<td>42% (75/178)</td>
<td>53% (205/384)</td>
</tr>
</tbody>
</table>
Discussion
Overall, subjects examined approximately two-thirds of the available relevant information. As predicted, they studied the base-rate more often on causal problems. Subjects also examined the contrapositive significantly more often on causal problems. Thus, on causal problems, subjects were more likely to attend to all the relevant information. It was predicted that subjects would be more inclined to seek out base-rates on problems for which they had prior expectations about the base-rates. This hypothesis was not supported by the data. Indeed, subjects were slightly (but not significantly) less likely to examine the base-rate on problems for which people typically have prior expectations. Prior expectations also did not affect the rate of seeking hit-rate and contrapositive information. Thus, prior expectations apparently did not affect information seeking.

The accuracy of subjects' final judgments was measured relative to Bayes' Theorem. Overall, the estimates produced by subjects departed dramatically from those predicted by Bayes' Theorem. The overall mean absolute value of the difference between subjects' estimates and those produced by Bayes' Theorem was 27.5%. Subjects were generally more accurate when there was a causal relationship described in the problem. This improved accuracy seems to have resulted from the increased use of the base-rate and the contrapositive.

Two types of gist representation were associated with differences in seeking base-rate information. Subjects expressing a correct representation of the base-rate examined the base-rate more than other subjects. As in the first experiment, subjects making the conversion error $P(A|B) = P(B|A)$ sought out the base-rate less often than other subjects. However, subjects making the $P(A\neg B) = P(\neg A|A)$ conversion error did not seek the base-rate less often. In the first experiment, it was not possible to distinguish between the effects of the two different kinds of conversion error. In the second experiment, despite the high degree of overlap between subjects making the two kinds of conversion error, only subjects committing a hit-rate conversion error examined the base-rate significantly less than other subjects. Thus, it appears that the conversion error effect stems from the representation of the hit-rate rather than the contrapositive.

With respect to seeking base-rates, the effects of a correct representation of the base-rate are independent of the effects of a conversion error. Overall, subjects expressing an appropriate representation of the base-rate and no conversion error examined the base-rate 76% of the time. By way of contrast, those subjects expressing an inappropriate representation of the base-rate and a conversion error examined the base-rate only 24% of the time. Thus, in understanding 'the base-rate fallacy' it is necessary to take into account the way people represent the hit-rate as well as the base-rate. Causal problems affected gist descriptions of the base-rate, but they did not affect gist descriptions of the hit-rate. So it appears that the locus of the causal relationship effect lies in people's representation of the base-rate.

EXPERIMENT 3

The relationship between conversion errors and base-rate neglect was established in the first two experiments by demonstrating the relationship between subjects' verbal descriptions of information items and their information-seeking behavior. The purpose of the third experiment was to provide direct corroborating experimental evidence for this effect. Toward this objective, a training paradigm was employed whereby some subjects received written and pictorial tutorials on the nature of the hit-rate and a control group received no such training. It was hypothesized that training about the nature of the hit-rate would lead to an increase in the seeking of base-rates and improved accuracy relative to Bayes' Theorem.
Method

Subjects
Forty-two undergraduate students in a required first-year course on learning from information technology served as subjects on a voluntary basis. Data were omitted for two subjects because they responded by hitting a single key on most problems, and appeared to be 'rushing' through the exercise rather than problem solving. None of these subjects had formal training in statistics.

Design and materials
Each subject received five conditional probability problems on an Apple Macintosh computer. Four problems were versions of the Ball and Jar problem (Appendix 1) and the Dormitory problem (Appendix 4) was last. Each problem presented subjects with the problem and menu of relevant information as in Experiments 1 and 2. Subjects were randomly assigned to either a training or a control condition. Subjects in the training condition received brief tutorials in the form of screens describing the meaning of \( P(A|B) \) after problems one, two, and three. The first tutorial consisted of a screen of text (see Exhibit 7), the second included a pictorial representation of the hit-rate concept (see Exhibit 8), and the third tutorial consisted of a repetition of both. The control group received no training.

Procedure
The procedure was identical to that used in Experiments 1 and 2 except that several subjects were run simultaneously at separate terminals in the same room, and each subject received an 'on-line walk through' of the procedure.

Results
After three brief on-life tutorials about the meaning of the hit-rate, significantly more subjects in the tutorial condition choose to examine base-rate information on Ball and Jar problem four than did control subjects, 83% versus 50%, \( \chi^2(1) = 4.84, p < 0.03 \). Unfortunately, subjects did not demonstrate transfer of training to the dormitory problem, \( \chi^2 < 1 \). Training subjects were also more accurate than control subjects on problem four relative to the mean absolute difference between their answers and those produced by Bayes' Theorem, 16.6 versus 28.3, \( F(1,39) = 4.35, p < 0.044 \). There were no other significant differences in accuracy or information-seeking behavior.

Feedback
The phrase of 'Of all the red balls drawn in this game, the percentage that came from the black jar' tells us how likely it is for a red ball to be drawn from the black jar as opposed to other jars. This gives us an indication of how 'rich' the black jar is in terms of red balls relative to other jars. In the language of hypothesis testing, this is the hit rate, or the probability that the jar is black given that we know that a red ball has been drawn. This does not directly tell us the probability of drawing a red ball from the black jar since even if this number is high, an equal or even higher percentage of white balls may have been drawn from the black jar as well.

Exhibit 7. Text tutorial explaining the meaning of the hit-rate on the Ball and Jar problem
Graphic Feedback: In the problem, You are at a carnival where games of chance are being played. One game requires you to draw a ball from one of several jars, each of which contains a number of balls that are either red or white. If you select a red ball, you win. On your turn you have to pick from the black jar. What is your estimate of the probability of drawing a red ball from the black jar?

Of all the red balls drawn in this game, the percentage that came from the black jar is 50%.

Exhibit 8. Graphic tutorial explaining the meaning of the hit-rate on the Ball and Jar problem

Discussion
The results of Experiment 3 confirm and reinforce an important conclusion from the first two experiments. Conversion errors in subjects’ reasoning about the hit-rate lead subjects to ignore base-rate information. When these misconceptions are reduced through educational interventions, subjects seek out the base-rate significantly more often. These results are consistent with those of Fong et al (1986), indicating that a little bit of statistical training can improve performance. Fong et al. (1986) found that people make use of intuitive, abstract inferential rules, and that formal training can increase the frequency and quality of reasoning with appropriate statistical heuristics. It appears that, although denominator neglect is a common simplification strategy, interventions may help subjects develop more appropriate gist representations. Unfortunately, the simple interventions devised for this study were not equal to the more difficult task of inducing positive transfer of training.

GENERAL DISCUSSION

A persistent finding in work on human judgment and decision making is that people frequently fail to utilize base-rates (Bar-Hillel, 1980; Kahneman and Tversky, 1972; Median and Bettger, 1991). Contrary to claims that the base-rate fallacy is a ‘myth’ (Koehler, 1993), the results of the three experiments described here are consistent with previous findings, indicating that people have difficulty integrating base-rates into their decision processes (Hamn, 1994). It is also important not to ‘oversell’ the magnitude
of base-rate neglect. Overall, almost one-half of the subjects failed to examine base-rate information—even when it was literally at their fingertips. These results help clarify the meaning of ‘the base-rate fallacy’ as more than a tendency to underestimate base-rates. However, more important than re-confirming errors in subjects’ decision making (or creating new labels such as ‘the contrapositive fallacy’), this research advances our understanding of information seeking on conditional probability problems and the representation of base-rates, hit-rates, and the contrapositive.

This research examined information seeking on Bayesian conditional probability problems by creating an on-line information-seeking environment. The experimental paradigm employed in most of the previous work on base-rates measured the extent to which probability estimates reflect the numeric values associated with base-rate and specific case information. Although lacking in ecological validity, the information-seeking paradigm employed in the current work allowed us to determine whether subjects fail to adequately weigh base-rate data or whether they fail to use them at all. Moreover, creating an information-seeking environment on the computer avoids creating unwarranted social expectations in subjects (Schwarz et al., 1991). Consistent with research on pseudodiagnosticity (Beyth-Marom and Fischhoff, 1983; Doherty et al., 1979), these results indicate that people seek only a portion of the available information, even when that information is immediately available and the costs of information seeking are practically nothing.

Unlike experiments employing static information displays, it was found that information seeking was not influenced by the presence of irrelevant information. As Reyna and Brainerd (in press b) note, correct reasoning ‘involves more than just recognizing the appropriate gist in problem information. It also involves inhibiting interference from irrelevant details, editing out irrelevant gists, knowing the relevant reasoning principles (e.g. proportionality), retrieving that principle in context, and correctly implementing that principle.’ Conditional probability problems, with their layers of nested and overlapping inclusion relationships, make for complicated mental bookkeeping, and ‘denominator neglect’ (which is, perhaps, unfortunately named since it may inadvertently imply a faulty attention process) is a common simplification strategy. As Reyna and Brainerd (in press b) note, ‘Predictable errors result from an effort to avoid the complexity of these inclusion relationships’.

These finding are consistent with the fuzzy-trace theory principle of output interference, and highlight the utility of ignoring information. Output interference indicates that generating responses interferes with reasoning by creating noise that degrades performance (Brainerd and Reyna, 1989, 1990a; Reyna and Brainerd, 1989). Output interference suggests serious debilitating consequences for processing irrelevant information, and explains why people are well served by seeking only a small amount of relevant information. While it is clearly inappropriate to ‘accept the null hypothesis’ that irrelevant information has no impact on information seeking (perhaps stronger manipulations would have an impact), it is impressive to note that in the first experiment subjects examined irrelevant information only 15 out of 432 possible times, and that in the second experiment subjects examined irrelevant information only twice out of a potential 576 opportunities.

It may be useful to think about information seeking in terms of a ‘filter’ such that, ideally, all relevant information is examined and all irrelevant information ignored. If such a filter is less than perfect, two kinds of errors in information seeking are possible: ignoring relevant information and seeking irrelevant information. Considering the difficulties caused by processing irrelevant information due to output interference, it may sometimes be more adaptive for people to ignore information—even at the risk of missing something important—than to seek information that is likely to be irrelevant.

For problems describing a causal relationship, subjects sought out the base-rate more often than for other problems, and were generally more accurate in their estimates. An examination of the subjects’ gist descriptions of the relevance of the base-rate revealed that a larger number of subjects appropriately described the relevance of the base-rate on these problems. Thus, it appears that base-rate neglect is
reduced on problems describing a causal relationship because these problems encourage the formation and utilization of appropriate gist representations of the base-rate.

In many contexts, developing useful gist entails representing causally relevant information. Thus, subjects have considerable experience extracting causally relevant information and incorporating it into their gist representations. With the two causal problems used in Experiment 2, it is easy to generate verbal gist representations of the base-rates without specific reference to 'base-rate' as a statistical concept or relative frequency. Consider the description of the base-rate on the Tennis problem: ‘Of all the matches that Jean has ever played, the percentage that Jean has won is 90%.’ A non-statistical verbal description of the gist of this statement might be ‘Jean is a good tennis player’. The description of the base-rate on the Test problem is: ‘Of all the students who take the test, the percentage who pass the test is 40%.’ Similarly, a non-statistical verbal description of the gist of this statement is ‘the test is hard’. Of course, it may be that for people schooled in statistical reasoning the gist of such statements on non-causal problems is something like ‘the base-rate is low’.

The finding that conversion errors and the correct representation of the base-rate independently affect the frequency of seeking base-rates indicates that ‘the base-rate fallacy’ is not a single phenomenon. Other researchers have interpreted confusion about the difference between \( P(A|B) \) and \( P(B|A) \) as a simple misreading of the problem statement. Macchi (1994, p. 2) argues that such confusion is a result of an ‘ambiguous transmission of information produced by the problem text’ and Koehler (1994, p. 8) further suggests that ‘we should think more about how to present probabilities in ways that minimize misunderstandings’. The fuzzy-trace principle of denominator neglect (Reyna, 1991) suggests that conversion errors are more than a simple misunderstanding of the problem text. From this perspective, conversion errors stem from difficulties in representing multiple, nested relationships, and a general preference for reasoning with gist-like representations of problems.

In the current work, the four problems did not differ reliably with respect to the frequency of conversion errors. However, it is reasonable to assume that problems drawn from other domains may differ in this respect. For example, Eddy (1982) found that approximately 95% of the physicians he interviewed exhibited a conversion error in their reasoning about the accuracy rate of an X-ray test. Conversely, other domains may yield problems for which conversion errors are rare. It may be that factors such as statistical training (Arkes, 1986; Fong et al., 1986; Nisbett et al., 1982) increase the use of base-rates by reducing the frequency of conversion errors.

The present experiments found that people often fail to seek out contrapositive information. Only a small minority of people (10%) apparently had an appropriate representation of the contrapositive. Thus, as in Beyth-Marom and Fischhoff’s (1983) research, it seems that even those people who did use the contrapositive may not have understood its true meaning. The erroneous belief that \( P(A|\neg B) = P(\neg B|A) \) appears to be widespread. For people who hold such an erroneous view, seeking contrapositive information may be more damaging to estimates of the conditional probability \( P(B|A) \) than not seeking the contrapositive at all. This helps explain why a positive test strategy (Fratianne and Cheng, 1991; Freedman, 1992; Klayman and Ha, 1987) can sometimes be more productive than a falsification strategy. This type of conversion error may also explain why falsification was counterproductive for some subjects on the Mynatt et al. (1978) simulated research task.

The results of this research are at odds with Lynch and Ofir’s (1989) claim that the base-rate fallacy is only an artifact of high case cue values. Their results are probably due to a confounding of case cue value and pragmatic case cue reliability. Pragmatically, ignoring the opinion of an inaccurate mechanic or witness is adaptive and appropriate. The materials used in the present work did not suffer from this confound, and the pervasive nature of base-rate neglect was demonstrated over three experiments and four diverse problems.

Over two decades of research on judgment and decision making have underscored the heuristic nature of decision making and reasoning. However, the 'heuristics and biases' approach to the study of
judgment and decision making has been criticized for producing a ‘laundry list’ of heuristics and biases that do not adequately describe empirical phenomena or underlying psychological processes. Fuzzy-trace theory offers a powerful theoretical framework for integrating research on specific heuristics. An important task that lies ahead is integrating well-researched heuristics into a unifying fuzzy-trace theoretical framework. Another worthy direction lies in the area of education. Of particular importance are efforts to develop quantitative intuition and facilitate the extraction of appropriate gist (Wolfe, 1993). The current work suggests that to teach people to avoid the base-rate fallacy, it is necessary to address the way they represent the hit-rate. Students are sometimes alerted to the base-rate fallacy when they are admonished, ‘When you hear hoof beats, think horses, not zebras’. Perhaps they must also be warned that ‘While horses make hoof beats, not all hoof beats are made by horses’.

APPENDIX 1: COMPUTER SCREEN PRESENTING THE BALL AND JAR PROBLEM STATEMENT AND INFORMATION MENU (NOT CAUSAL, NO EXPECTATIONS)

YOU ARE AT A CARNIVAL WHERE GAMES OF CHANCE ARE BEING PLAYED. ONE GAME REQUIRES YOU TO DRAW A BALL FROM ONE OF SEVERAL JARS, EACH OF WHICH CONTAINS A NUMBER OF BALLS THAT ARE EITHER RED OR WHITE. IF YOU SELECT A RED BALL, YOU WIN. ON YOUR TURN, YOU HAVE TO PICK FROM THE BLACK JAR. WHAT IS YOUR ESTIMATE OF THE PROBABILITY OF DRAWING A RED BALL FROM THE BLACK JAR?

1) OF ALL THE RED BALLS DRAWN IN THIS GAME, THE PERCENTAGE THAT CAME FROM THE BLACK JAR.
2) THE PERCENTAGE OF BALLS DRAWN FROM ALL OF THE JARS IN THIS GAME THAT WERE DRAWN BY STUDENTS.
3) OF ALL THE WHITE BALLS DRAWN IN THIS GAME, THE PERCENTAGE THAT CAME FROM THE BLACK JAR.
4) OF ALL THE RED BALLS DRAWN IN THIS GAME, THE PERCENTAGE THAT WERE DRAWN BY WOMEN.
5) THE PERCENTAGE OF BALLS DRAWN FROM ALL OF THE JARS IN THIS GAME THAT ARE RED.
6) OF ALL THE WHITE BALLS DRAWN IN THIS GAME, THE PERCENTAGE THAT WERE DRAWN BY MEN.

APPENDIX 2: COMPUTER SCREEN PRESENTING THE TENNIS MATCH PROBLEM STATEMENT AND INFORMATION MENU (CAUSAL RELATIONSHIP, NO EXPECTATIONS)

YOUR FRIEND JEAN IS PLAYING IN A TENNIS MATCH. THE MATCH IS BEING PLAYED ON A CLAY TENNIS COURT, AND YOU’VE NEVER SEEN JEAN PLAY ON A CLAY SURFACE. YOU WOULD LIKE TO ESTIMATE THE CHANCES OF JEAN WINNING THIS MATCH. WHAT IS YOUR ESTIMATE OF THE PROBABILITY OF JEAN WINNING THIS TENNIS MATCH ON A CLAY COURT?

1) OF ALL THE MATCHES THAT JEAN HAS EVER PLAYED, THE PERCENTAGE THAT JEAN HAS WON.
2) OF ALL THE MATCHES THAT JEAN HAS EVER PLAYED, THE PERCENTAGE THAT WERE PLAYED IN SUMMER.
3) OF ALL THE MATCHES THAT JEAN HAS EVER WON, THE PERCENTAGE THAT WERE PLAYED ON A CLAY COURT.
4) OF ALL THE MATCHES THAT JEAN HAS EVER WON, THE PERCENTAGE OF TIMES SHE WAS WEARING WHITE.
5) OF ALL THE MATCHES THAT JEAN HAS EVER LOST, THE PERCENTAGE THAT WERE PLAYED ON A CLAY COURT.
6) OF ALL THE MATCHES THAT JEAN HAS EVER LOST, THE PERCENTAGE OF TIMES SHE WAS USING AN ORANGE BALL.

APPENDIX 3: COMPUTER SCREEN PRESENTING THE DORMITORY PROBLEM STATEMENT AND INFORMATION MENU (NOT CAUSALLY RELATED, PRIOR EXPECTATIONS)

JOHN IS A STUDENT AT A SMALL COLLEGE WHO IS INTERESTED IN FORMING A PSYCHOLOGY DEPARTMENT INTRAMURAL BASKETBALL TEAM. ANYONE CAN PLAY ON THE TEAM AS LONG AS THEY HAVE HAD AT LEAST ONE PSYCHOLOGY COURSE. JOHN IS DECIDING WHETHER HE SHOULD TRY TO RECRUIT PLAYERS FROM HIS DORM. WHAT IS YOUR ESTIMATE OF THE PERCENTAGE OF STUDENTS IN JOHN'S DORM WHO HAVE TAKEN A PSYCHOLOGY COURSE?

1) OF ALL THE STUDENTS WHO HAVE HAD A PSYCHOLOGY COURSE, THE PERCENTAGE WHO LIVE IN JOHN'S DORM.
2) OF ALL THE STUDENTS IN JOHN'S COLLEGE, THE PERCENTAGE WHO HAVE SEEN A HOCKEY GAME.
3) OF ALL THE STUDENTS WHO HAVE NEVER HAD A PSYCHOLOGY COURSE, THE PERCENTAGE WHO LIVE IN JOHN'S DORM.
4) OF ALL THE STUDENTS WHO HAVE HAD A PSYCHOLOGY COURSE, THE PERCENTAGE WHO ARE SOPHOMORES.
5) OF ALL THE STUDENTS IN JOHN'S COLLEGE, THE PERCENTAGE WHO HAVE HAD A PSYCHOLOGY COURSE.
6) OF ALL THE STUDENTS WHO HAVE NEVER HAD A PSYCHOLOGY COURSE, THE PERCENTAGE WHO HAVE HAD PHILOSOPHY.

APPENDIX 4: COMPUTER SCREEN PRESENTING THE PSYCHOLOGY TEST PROBLEM STATEMENT AND INFORMATION MENU (CAUSALLY RELATED PRIOR EXPECTATIONS)

THE PSYCHOLOGY DEPARTMENT OF A SMALL COLLEGE REQUIRES ITS STUDENTS TO PASS A PSYCHOLOGY TEST IF THEY ARE TO GRADUATE FROM THE DEPARTMENT WITH HONORS. A STUDENT WITH AN 'A' AVERAGE IS NOW PREPARING TO TAKE THE TEST. WHAT IS YOUR ESTIMATE OF THE PROBABILITY THAT THIS STUDENT WILL PASS THE TEST?

1) OF ALL THE STUDENTS WHO PASS THE TEST, THE PERCENTAGE WHO ARE 'A' STUDENTS.
2) OF ALL THE STUDENTS WHO TAKE THE TEST, THE PERCENTAGE OF STUDENTS WITH A MINOR IN BIOLOGY.
3) OF ALL THE STUDENTS WHO FAIL THE TEST, THE PERCENTAGE WHO ARE ‘A’ STUDENTS.
4) OF ALL THE STUDENTS WHO PASS THE TEST, THE PERCENTAGE WHO ARE TALL.
5) OF ALL THE STUDENTS WHO TAKE THE TEST, THE PERCENTAGE OF STUDENTS WHO PASS THE TEST.
6) OF ALL THE STUDENTS WHO FAIL THE TEST, THE PERCENTAGE WHO ARE SHORT.

REFERENCES

Hamm, R. M. ‘Underweighing of base-rate information reflects important difficulties people have with probabilistic inference’, *PSYCOLOQUY*, 5 (1994).


Klar, Y. 'Linking structures and sensitivity to judgment relevant information in statistical and logical reasoning tasks', Journal of Personality and Social Psychology, 59 (1990), 841–858.


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


Wilkins, M. C. 'The effect of changed material on ability to do formal syllogistic reasoning', Archives of Psychology, 102 (1928), 83.

Author's biography:
Christopher R. Wolfe is Director of Quantitative Reasoning and Instructional Computing in the School of Interdisciplinary Studies at Miami University. He is an assistant professor of Interdisciplinary Studies and an affiliate of the Department of Psychology. Dr. Wolfe holds a doctorate in cognitive psychology from the University of Pittsburgh.

Author's address:
Christopher R. Wolfe, School of Interdisciplinary Studies, Miami University, Oxford, OH 45056, USA.