JUDGMENTAL BIAS AND THE TWO-EXPERIMENT PROBLEM

by

Donald Sharpe

A thesis
presented to the University of Manitoba
in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy

Winnipeg, Manitoba

(c) Donald Sharpe, 1992

National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-77752-4

Canada

JUDGMENTAL BIAS AND THE TWO-EXPERIMENT PROBLEM

BY

DONALD SHARPE

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

© 1992

Permission has been granted to the LIBRARY OF THE UNIVERSITY OF MANITOBA to lend or sell copies of this thesis, to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film, and UNIVERSITY MICROFILMS to publish an abstract of this thesis.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's permission.

ABSTRACT

Tversky and Kahneman (1974) proposed that individuals invoke heuristics to make judgments, and concluded that these short-cuts often lead to error. Although widely accepted, this model of the error-prone decision maker has been criticized as more parsimoniously explained as a misinterpretation of subjects' behavior. If some subjects do not interpret judgment tasks in the manner intended by the experimenter, what Orne (1962) called the two-experiment problem, the subjects' behavior is misconstrued.

Five experiments explored subjects' perceptions of classic judgment problems within Orne's framework. In Experiment 1, subjects' responses to judgment problems deviated significantly from the "correct" solutions derived from the normative model. Assessments of their perceptions of the experimental task were found to be related to their deviations scores, but inconsistent with the dictates of the normative model. In a second experiment, subjects' correct identification of base-rate and likelihood information were not found to result in smaller deviations from normative solutions. In a third experiment, modifications of the problems that had been previously shown to reduce deviation scores were found to be effective and to lead subjects to interpretations of the problems more consistent with that intended by experimenters. Smaller deviation scores, however, were not attributed to adopting more normative

interpretations for the task. A fourth experiment sought to manipulate subjects' pre-experimental perceptions of the judgment task, and a fifth experiment permitted subjects to discuss each problem prior to offering a solution. Deviation scores were not affected by either manipulation. The results from the five experiments were discussed in terms of subjects' holding a different interpretation of the task than intended by experiments and implied by the normative model, and their responses to the problems being consistent with this interpretation.

ACKNOWLEDGEMENTS

Although only the author of this dissertation was awarded the doctoral degree, many individuals contributed to this project. Firstly, I wish to thank John Adair for his knowledge, guidance, and patience. When I was deciding whether to go to graduate school at the University of Manitoba, I asked someone for advise. When I said my advisor would be John Adair, she replied "Go!". I've never received any better counsel. Second, I wish to thank Jim Forest, Cam Huynh, Bill Notz, and Ted Wright for serving on my dissertation committee. Their participation in the project is much appreciated.

On a personal level, I wish to thank my friends for their support. In particular, I would like to thank Steve and Louise Cook, Darl Ramsum, Kevin Brown, Debbie Olsen, Marcia Pollock, Linda Temple, Lorna Jakobson, David, Ann, Mark, Thomas, and Andrew Patton, Jocelyn Proulx and Dwayne Dehanyk, Kate Tunna, Deborah Gilman, Kathryn Ritchot, Shannon Koshelanyk, Norah Vincent, Diane Forest, Gerry Sande, Ward Struthers, Trevor Markesteyn and Mary Borys, Susan Nadon, Susan Bruce, Cindy Cain, Arlen Nimchuk, Andy Lubusko, Tim Osachuk, Neharika Vohra, and Patti Whyte. I also want to thank Reg, Robert and Tracy Sharpe, and Augie. I dedicate this dissertation to my father, Arthur Sharpe, who I think would have been proud of me, and to my mother, Lillian Sharpe, who I love very much.

TABLE OF CONTENTS

Page

ABSTRACT.....i

ACKNOWLEDGEMENTS.....iii

TABLE OF CONTENTS.....iv

INTRODUCTION.....1

Judgmental Heuristics.....4

Representativeness.....7

Bias in the Social Judgment Domain.....22

Fundamental Attribution Error.....24

Humans as Irrational Decision Makers.....26

Limitations to Normative Models.....29

Ecological Validity.....32

Context.....34

Subjects' Interpretations.....36

Two-Experiment Problem.....40

Statement of the Problem.....45

EXPERIMENT 1.....55

Method.....60

Results.....62

Discussion.....80

EXPERIMENT 2.....84

Method.....86

Results.....87

Discussion.....91

EXPERIMENT 3.....93

Method.....95
Results.....100
Discusssion.....119
Experiment 4.....123
Method.....126
Results.....129
Discussion.....138
Experiment 5.....142
Method.....146
Results.....146
Discussion.....156
GENERAL DISCUSSION.....158
REFERENCES.....166
APPENDIX A.....188
APPENDIX B.....208
APPENDIX C.....218

JUDGMENTAL BIAS AND THE TWO-EXPERIMENT PROBLEM

What is a rational decision? Although both the classic and contemporary literature have portrayed people as rational, defining the nature of rational thought is difficult (Kyburg, 1983). One perspective on rationality, which appeared in the Renaissance and was later reintroduced by von Neumann and Morgenstern (1953), called for the application of "suitable mathematical methods" (p. 9) to explain how decisions should be made. According to utility theory (von Neumann & Morgenstern, 1953), decisions should be made by weighing the potential consequences of an outcome against assets now held. If this evaluation follows the dictates of probability theory, then the decision that results is rational. On the other hand, if this evaluation is based on habit, tradition, social desirability, or religious principles, then the decision that results is not rational (Dawes, 1988).

One form of probability theory has been most closely associated with rational decision making. Bayes' theorem can be expressed as

$$p(H_1/D) = p(H_1) \times p(D/H_1)/p(D)$$

$$\text{with } p(D) = p(H_1) \times p(D/H_1) + p(H_2) \times p(D/H_2)$$

The left side of the first equation is the posterior probability, the probability that the first of a set of hypotheses is true given the data. The posterior probability is the product of two elements. The first is

the base-rate, the prior probability that the first hypothesis is true before the data are given. The second is the likelihood ratio, individuating information, or the impact of the data if the first hypothesis is true (Gigerenzer & Murray, 1987).

Many studies in the 1950's and 60's tested Bayes' theorem as a model of human decision making (e.g., Peterson & Beach, 1967). Subjects were presented with problems taken directly from probability textbooks. Urn-and-balls problems, still commonly found in introductory statistics textbooks (e.g., Gravetter & Wallnau, 1988), were the favorites of researchers of the era. One such problem is reproduced below:

> Two urns are filled with a large number of poker chips. The first urn contains 70% red chips and 30% blue. The second contains 70% blue chips and 30% red. The experimenter flips a fair coin to select one of the two urns, so the prior probability for each urn is .50. He then draws a succession of chips from the selected urn. Suppose the sample contains eight red and four blue chips. What is your revised probability that the selected urn is the predominantly red one? (Peterson & Beach, 1967, p. 32)

According to Bayes' theorem, .97 is the revised or posterior probability that the selected urn is the

predominantly red one. [1] The posterior probability is obtained by combining the base-rate of .50 and the likelihood information (the sample of eight red and four blue chips). The modal response given by subjects is .75. The difference between the posterior probability derived from Bayes' theorem and subjects' modal response was seen to result from insufficient revision from the base-rate.

The conclusion of that literature was human decision making is rational. Human decision making follows the principles of probability theory, but human decision makers are conservative Bayesians (Peterson & Beach, 1967) --

---

[1] The prior probabilities of the hypotheses H_1 (70% red chips and 30% blue chips) and H_2 (70% blue chips and 30% red chips) are $p(H_1)$ and $p(H_2)$. The likelihood of the data D if H_1 and H_2 are true are $p(D/H_1)$ and $p(D/H_2)$, respectively. Therefore

$$p(D/H_1) = (n/x)p^{x1} \quad (1-p_1)^{n-x}$$

$$= (12/8) \ .7^8 \times (.3)^4$$

$$= .231$$

and

$$p(D/H_2) = (12/8) \ .3^8 \times (.7)^4$$

$$= .008$$

therefore the posterior probability is equal to

$$p(H_1/D) = .5 \times .231/(.5 \times .231 + .5 \times .008)$$

$$= .967$$

(Gigerenzer & Murray, 1987, p. 148-149).

failing to adjust sufficiently from the base-rate value as a function of the likelihood information. What is the source of this error? Although conservatism has "roots in the fundamental aspects of subjects' understanding and use of information" (p.33), as a decision maker, "man gambles well. He survives and prospers while using the fallible information to infer the states of his uncertain environment and to predict future events" (p. 29, emphasis added). Thus, the literature of the 1950's and 60's concluded that irrationality in human decision making was largely the product of factors external to the decision maker (Wright & Ayton, 1990).

Judgmental Heuristics

In the early 1970's, Daniel Kahneman and Amos Tversky (1972, 1973; Tversky & Kahneman, 1974) offered a radically different perspective on human decision making. The problems they presented subjects were derived from probability theory but more closely approximated real-world judgments than urn-and-balls style questions. One widely-cited demonstration is the following:

> Imagine that the U.S. is preparing for an outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Scientific estimates of the consequences of the program are as follows:

> If program A is adopted, 200 people will be saved. If program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.
>
> Which of the two programs would you favor?
>
> (Kahneman & Tversky, 1984, p. 343)

Regardless of the program adopted, an outcome of saving 200 people is equally likely. Nevertheless, about three-quarters of subjects chose program A.

The same problem with different alternatives -- "If program C is adopted, 400 people will die. If program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die" (p. 343) has an equally likely outcome of saving 200 people regardless of which program is adopted. Nevertheless, about three-quarters of subjects presented with these alternatives chose program D. Whether framed as people living or dying, programs A and C, and programs B and D, are equivalent. Rationality demands a choice be made on the basis of its consequences. When the consequences are the same, neither program should be favored (Tversky & Kahneman, 1981).

On the basis of such demonstrations, Kahneman and Tversky developed an influential research program that resulted in two significant conclusions. Their first conclusion was that probability theory, specifically Bayes'

theorem, is not an appropriate representation of human decision making. Kahneman and Tversky (1972) determined "man is apparently not a conservative Bayesian: he is not Bayesian at all" (p. 450). Kahneman and Tversky proposed short-cut strategies or heuristics to replace probability theory as the model for human decision making. "People rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (Tversky & Kahneman, 1974, p. 1124). Reducing stimulus complexity was central to many models of cognitive processing popular at that time (e.g., Kahneman, 1973; Treisman, 1969). Although conceding heuristics are sometimes useful, Tversky and Kahneman (1974) felt these shortcuts often result in "severe and systematic errors" (p. 1124).

One heuristic is availability. Tversky and Kahneman (1974) suggested "there are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind" (p. 1127). Error arises when other significant factors are ignored in favor of the instance most readily available from memory. Examples of faulty judgment attributed to an availability heuristic include publicized causes of death judged as more frequent, e.g., homicide versus diabetes (Fischhoff, Slovic, & Lichtenstein, 1977), and estimating the number of seven-letter words of

the form _____ing as more common than the number of seven-letter words of the form _____n_ (Tversky & Kahneman, 1974).

A second heuristic identified is anchoring and adjustment. The first instance of an object or an event will serve as an anchor for later judgments (Tversky & Kahneman, 1974), as an "orienting point for estimating probabilities or outcomes" (Dawes, 1988, p. 121). Tversky and Kahneman (1974) proposed decision makers often underadjust from this anchor when making probability estimates. In one demonstration (Tversky & Kahneman, 1974), subjects estimated whether the percentage of African countries in the United Nations was greater or less than a number that came up on a rigged spinning wheel. The correct response was 35 percent. If the number 10 appeared on the wheel, the median estimate made by subjects was 25 percent. On the other hand, if the number 65 appeared on the wheel, subjects' median estimate was 45 percent.

Representativeness

A third heuristic is representativeness (Kahneman & Tversky, 1972; Tversky & Kahneman, 1974). When asked whether an object labelled A is part of a class of objects labelled B, Tversky and Kahneman (1974) hypothesized "probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B" (p. 1124).

Of the heuristics originally identified by Kahneman and

Tversky, the representativeness heuristic has generated by far the most interest because it has served as an explanation for some of the most compelling biases in human judgment. The base-rate fallacy is a bias attributed to the representativeness heuristic. If a man is described as shy, helpful and tidy, one is more likely to infer that he is a librarian rather than a farmer, failing to consider there are many more farmers than librarians (Tversky & Kahneman, 1974). Insensitivity to sample size has also been attributed to representative thinking. An event with a low probability, such as the average height of a sample of men being greater than six feet, is judged to be equally likely regardless of the size of the sample. Misperceptions of chance and regression to the mean (Tversky & Kahneman, 1974) are other errors in judgment that have been attributed, in part, to the representativeness heuristic.

Tom. Investigation of the representativeness heuristic have been undertaken through several classic demonstrations. One such demonstration required subjects to determine the occupation of Tom (Kahneman & Tversky, 1973). Subjects in a Base-Rate condition were asked to estimate the percentage of graduate students in each of nine areas of study. Subjects in a Similarity condition were presented with the following description of a high-school student and asked to judge how similar Tom was to a typical graduate student in each of the nine areas of study:

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense. (p. 238)

Subjects in a Prediction condition were provided with the same description, but were told it was written when Tom was in high school by a psychologist who based his description on projective tests. These subjects were also asked to describe how likely it was that Tom was now a graduate student in each of the nine areas of study.

Rankings of how similar Tom was to the typical graduate student in each area of study were positively associated for Prediction and Similarity condition subjects, but were negatively associated with the rankings of Base-Rate condition subjects. Thus, Prediction and Similarity condition subjects appeared to ignore the base-rate information, and to rely solely on the personality sketch in making their rankings. Prediction condition subjects also

violated a "fundamental rule of statistical prediction ... that expected accuracy controls the relative weights assigned to specific evidence and to prior information" (Kahneman & Tversky, 1973, p. 239). Prediction condition subjects should have judged the accuracy of the personality description to have been low because the description of Tom W. was written when he was in high school and was based on projective tests. From Kahneman and Tversky's interpretation of the problem, a 'rational' decision maker should have disregarded the personality description, and have relied exclusively on the base-rate.

Jack Problem. The Jack problem is similar to the Tom problem, but provides a stricter test of the representativeness hypothesis (Kahneman & Tversky, 1973). Subjects were told that descriptions of 30 engineers and 70 lawyers had been written by psychologists based on interviews and personality tests. Subjects were told that "experts" had been able to accurately assign probabilities as to whether an individual was an engineer or a lawyer based on these descriptions. Five personality descriptions followed, one being of an individual named Jack:

> Jack is a 45 year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home

carpentry, sailing, and mathematical puzzles. The probability that Jack is one of the 30 engineers in the sample of 100 is _____%. (p. 241)

Following the same reasoning as offered for the Tom W. problem, the probability judgment for Jack's occupation should have been the base-rate.[2] Regardless of whether the base-rate specified for Jack was 30 engineers and 70 lawyers or 70 engineers and 30 lawyers, however, probability judgments of about 50 percent were made.[3] The same result was found for a description that was uninformative as to the occupation of the target individual. Only when no description of the target individual was given did ratings

---

[2]The formula for calculating the relationship between the 70/30 and 30/70 base-rate conditions is

$$\frac{p(E/HE)}{1-p(E/HE)} \Big/ \frac{p(E/LE)}{1-p(E/LE)} = 5.44$$

$$\frac{.70}{1-.70} \Big/ \frac{.30}{1-.30} = 5.44$$

(Wells & Harvey, 1978, p. 486)

[3]Comparing the two versions of the Jack problem, Kahneman and Tversky reported the mean posterior probability for the low-engineer group to be 50 percent, for the high-engineer group to be 55 percent. Kahneman and Tversky dismissed this statistically significant difference and instead examined the median judged probability against a line derived from Bayes' theorem. Gigerenzer and Murray (1987) concur that null hypothesis testing was inappropriate in this instance since they regard there to be two competing predictions, and thus no null hypothesis to be rejected.

approximate the base-rate.

Several studies have replicated Kahneman and Tversky's original finding. Most of these studies have manipulated either the base-rate or the description of the target. Manipulating the base-rate values has been consistently shown to have little impact on ratings (Fischhoff & Bar-Hillel, 1984a). An exception is a study by Carroll and Siegler (1977) who constructed twenty personality descriptions said to have been drawn from a population of either twenty or four-hundred individuals. Subjects made judgments much closer to the base-rate when the sample to population ratio was 20/20 rather than 20/400, suggesting that base-rate information may be used if the sample-to-population ratio is small.

Studies that have varied the diagnostic value of the target descriptions have produced mixed results. Fischhoff and Bar-Hillel (1984a) classified Kahneman and Tversky's engineer/lawyer scenario as a high-diagnostic context, because the description fit the stereotype for one of the two occupations. A low-diagnostic context is the case in which the two target occupations are similar, such as engineer vs physicist or business executive vs lawyer. Although judgments accurately reflected base rates in some conditions, base-rate information generally was ignored by subjects in both high- and low-diagnostic contexts (Fischhoff & Bar-Hillel, 1984a).

Ginosar and Trope (1980) reasoned that if the description of the target individual includes information conflicting with membership in a particular occupation (e.g., an engineer interested in political and social issues), base rates might have some greater impact. Compared to consistent individuating information about a target individual, inconsistent individuating information led to judgments that deviated less from the base-rate.

Gigerenzer, Hell and Blank (1988) found greater reliance on the base-rate with actual descriptions of real individuals employed in selected occupations, and when random sampling of the descriptions was observed by subjects. Contrary to Ginosar and Trope (1980), but consistent with Kahneman and Tversky (1973), Gigerenzer, et al., (1988) found a non-diagnostic description resulted in the base-rate being ignored. They concluded that whether a non-diagnostic description preceded or followed diagnostic descriptions explained this discrepancy.

The setting or context for the presentation of the problem also affects the use of base-rate information. Zukier and Pepitone (1984) presented an engineer/lawyer problem as part of either a scientific investigation (a base-rate appropriate setting) or a clinical job counsellor training study. Judgments more closely approximated the base-rate in the context of a scientific investigation -- more so for a non-diagnostic description, less so for the

stereotypical description. Similar results were obtained in a second study that compared medical school residents and first-year medical students. Zukier and Pepitone reasoned that medical school residents had been indoctrinated into a medical orientation, but that first-year medical students still retained a scientific orientation from their undergraduate training in the sciences.

Cab Problem. The Cab problem is similar in appearance to the urn-and-balls problems. One version of this problem:

> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (i) 85% of the cabs in the city are Green and 15% are Blue.
>
> (ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.
>
> Question: What is the probability that the cab involved in the accident was Blue rather than Green? (Tversky & Kahneman, 1982a, p. 156)

Expressed in terms of Bayes' theorem, the base rates

are 85 and 15 percent, the likelihood values are 80 and 20 percent. The resulting probability is .41 that the cab was Blue given the report of "Blue" by the witness.[4] Tversky and Kahneman reported subjects' modal response to be .80, the same value as that given for the accuracy of the witness, and suggested the base-rate had been ignored in favor of the likelihood value.

Subsequent studies have shown no effect for varying elements of the problem such as majority colour, the cab belonging to the larger company, or the order of presentation of base-rate and likelihood information (Bar-Hillel, 1980; Borgida & Brekke, 1981; Pollard & Evans, 1983; but see also Krosnick, Li & Lehman, 1990). Research has focused on manipulating the base-rate and likelihood

---

[4] The solution to the cab problem is as follows:

$$\frac{p(H/D)}{p(H/\sim D)} = \frac{p(D/H) \times p(H)}{p(D/\sim H) \times p(\sim H)} = \frac{.8 \times .15}{.2 \times .85}$$

$$= \frac{.12}{.17} = \frac{p(D)}{p(\sim D)}$$

$$p(H/D) = \frac{p(D)}{p(D) + p(\sim D)} = \frac{.12}{.12 + .17} = .41$$

(Tversky & Kahneman, 1982a, pp. 157-158. For a more intuitive explanation of the Bayesian solution, see Pollard & Evans, 1983).

information, and these studies have yielded mixed results. In a within-subject design, Fischhoff, Slovic and Lichtenstein (1979) presented a variety of base rates to subjects. A small number of subjects responded .80 (the likelihood value), regardless of the base-rate given. Another small group of subjects responded consistently with whatever the base-rate value was. Most subjects appeared sensitive to change in the base rates, although their probability judgments were often underadjusted from the likelihood value (see Scholz, 1987).

Tversky and Kahneman (1982a), in a modified replication of the Cab problem, sought to make the base-rate appear causal (see Ajzen, 1977). The base-rate statement in the original version of the Cab problem was replaced by "Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs" (p. 157). Tversky and Kahneman argued this change leads to a causal attribution that Green cab drivers are less cautious than Blue cab drivers. The median response to this causal version of the problem was .60, and the difference between subjects' median and modal response (.60) from the original demonstration (.80) was attributed to subjects more appropriately integrating base rates.

Varying the likelihood information, specifically the accuracy of the witness, has also been shown to effect subjects' probability judgments (e.g., Lyon & Slovic, 1976).

As the witness became less accurate, probability judgments decrease in value, and subjects' confidence in their own probability judgment decreases (Hinsz, Tinsdale, Nagao, Davis & Robertson, 1988).

Tversky and Kahneman (1982a) removed the information about the witness from the problem, and found subjects then responded with the base-rate. Bar-Hillel (1980) replaced the witness with actuarial information:

> The police investigation discovered that in the neighbourhood in which the accident occurred, which is nearer to the [Blue] Cab company headquarters than to the [Green] Cab company, 80% of all taxis are [Blue], and 20% are [Green]. (p. 226) [5]

The modal response was 80 percent, the same value as reported for the original presentation of the problem. Another version had two witnesses. The first witness said the cab was blue, and was found to be correct 80 percent of the time. A second witness said the cab was green, and was found to be correct 70 percent of the time. The modal response for the probability the cab was blue fell to 55 percent. Finally, Bar-Hillel obtained a median response of 48 percent with the following changes to the problem:

---

[5]The color of the two cabs is reversed in this quotation and subsequent discussion of Bar-Hillel's studies in order to be consistent with Kahneman and Tversky's (1982a) version of the problem discussed earlier.

> Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are [Green], and 15% are [Blue]. A cab was involved in a hit-and-run accident at night, in which a pedestrian was run down. The wounded pedestrian later testified that though he did not see the color of the cab due to the bad visibility conditions that night, he remembers hearing the sound of an intercom coming through the cab window. The police investigation discovered that intercoms are installed in 80% of the [Blue] cabs, and in 20% of the [Green] cabs. (p. 228)

Bar-Hillel argues this version of the problem succeeds because it reduces the perceived relevance of the likelihood information.

Lynch and Ofir (1989) varied both the witness accuracy (20 or 80 percent) and the base-rate (15 or 85 percent), with both base-rate and witness accuracy or only one of the two sources of information being presented to subjects. The accuracy of the witness was inconsequential when paired with a base-rate of 85 percent, but was significant when the base-rate was 15 percent. Conversely, the base-rate had little effect when the witness accuracy was 80 percent, but had a significant effect when the witness accuracy was 20 percent -- although knowing that a witness is correct only

20% of the time has the same diagnostic value as knowing that a witness is correct 80% of the time (Tversky and Kahneman, 1982a). Lynch and Ofir (1989) proposed that if the cab problem had originally been presented by Kahneman and Tversky with a high value for the base-rate and the accuracy of the witness had been varied, researchers might have identified a 'case-rate' rather than base-rate fallacy.

The Cab problem highlights the complex relationship between the representativeness heuristic and the so-called base-rate fallacy. Tversky and Kahneman (1974) had originally attributed many judgmental phenomena, including the base-rate fallacy, to the representativeness heuristic. Bar-Hillel (1980), among others, has pointed out base-rate neglect can occur without representativeness. For example, it is not apparent how the resemblance between a sample and a population is relevant to solving the Cab problem. Bar-Hillel believes representativeness may be only one of a number of factors that influence the perceived relevance of baserates.

In response to these concerns, Tversky and Kahneman (1982b) differentiated judgments by representativeness from judgments of representativeness. The former is the heuristic, the latter the case in which subjects expect samples to be representative of their population. Judgment by representativeness was further differentiated by Tversky and Kahneman into several classes that Gigerenzer and Murray

(1987) conclude can be combined into overreliance on the likelihood information and ignoring the base-rate. Whatever the specific nature of the relationship between the base-rate fallacy and the representativeness heuristic, the former appears to serve as a better explanation for some demonstrations of judgmental bias (e.g., the Cab problem), the latter as a better explanation for other demonstrations (e.g., the Jack problem).

Linda. The Linda problem relates to the extension, conjunction or compound-probability fallacy (Tversky & Kahneman, 1982b, 1983). Although attributed by Tversky and Kahneman (1982b) to a representativeness heuristic, this bias shares some relationship to the availability heuristic (Dawes, 1988). The Linda problem is as follows:

> Linda is a 31 year old, single, outspoken, and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.
>
> Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable...
>
> Linda is active in the feminist movement (F) ...
>
> Linda is a bank teller (T) ...
>
> Linda is a bank teller and is active in the feminist movement. (T & F) (Tversky & Kahneman,

1982b, p. 93)

Only three of the eight options provided to subjects are reproduced above. Subjects in between-group conditions responded either to the compound (T & F) or to the two simple targets (F) and (T). Subjects in a within-group condition compared the compound and simple targets directly. Across all conditions, most subjects judged Linda to be more likely to be both a bank teller and a feminist than a bank teller. This outcome appears to violate the conjunction rule, that the combination of two elements cannot exceed the probability of any one of the constituent elements. Dawes (1988) suggests that "we can imagine her becoming a feminist bank teller, but it is hard to imagine her as merely a (stereotyped) bank teller, even though feminism was not mentioned in the description of Linda" (p. 130).

If the statement "Linda is a bank teller" is replaced by "Linda is a bank teller whether or not she is active in the feminist movement", which clarifies the relationship between bank teller and feminist, the bias is still found for more than half of subjects (Agnoli & Krantz, 1989; Morier & Borgida, 1984; Tversky & Kahneman, 1982b). Although subjects with training in statistics or logic do no better than naive subjects on the problem (Tversky & Kahneman, 1982b), the former are quicker to recognize their error if reminded of the conjunction rule (Agnoli & Krantz, 1989; Morier & Borgida, 1984; Tversky & Kahneman, 1982b).

Agnoli and Krantz (1989) reduced the frequency of violations of the conjunction rule, but only after providing extensive training to subjects.

Leddo, Abelson and Gross (1984) found the conjunction fallacy in several contexts, but also reported fewer subjects were biased when the action an actor tried to engage in was unsuccessful -- Fred did not stop at a Italian restaurant to eat. Fred was not hungry and he disliked Italian foods. Carlson and Yates (1989) found the frequency of disjunction errors, the violation of the principle that the disjunction of two events must be at least as likely as the probability of occurrence of each of the constituent events, to be comparable to the frequency of conjunction errors reported by Tversky and Kahneman (1983). Disjunction errors are of interest because it is difficult to understand how Linda could be representative of "either a feminist or a bank teller" (Reeves & Lockhart, 1990; emphasis added).

Bias in the Social Judgment Domain

Research into bias in probabilistic judgment has been influential, widely cited, and linked to a parallel stream of research investigating bias in social judgment. Kelley (1972) hypothesized the "naive psychologist" tests causal hypotheses through a process akin to analysis of variance. In order to determine the cause of their own behavior or the behavior of another person (the actor), subjects consider how other individuals have behaved in the same situation

(consensus), how the actor has behaved on other occasions (consistency), and how the actor has behaved toward other persons (distinctiveness). The combination of these dimensions (e.g., low consensus, low distinctiveness, high consistency), reveals the attribution to the cause of the event that should be made (e.g., the actor). In the same way that probability theory and Bayes' theorem provide the 'correct' probability judgment, attribution theory and the ANOVA cube provide the 'correct' causal attribution to be made by the subject (Kruglanski & Ajzen, 1983).

Similar to the descriptive function served by probability theory in the 1950's and 1960's, attribution theory was originally proposed as a representation of how human decisions are made (e.g., Heider, 1958). Funder (1987) argues that Kelley's ANOVA cube was "based on the working assumption that people are perfectly 'rational' processors of information" (p. 77). It soon became obvious, however, that individuals deviate from both probabilistic and attributional models (Ross, 1978). Consistent with the normative function probability theory currently serves in cognitive judgment research, attribution theory is now offered as the standard by which fallible social judgments are assessed (Gigerenzer & Murray, 1987). Thus, attributional models "came to be viewed less as theories of how people do make judgments and more as standards prescribing how people should make judgments" (Funder, 1987,

p. 77).

The close relationship between attribution theory and judgmental heuristics, specifically representativeness, is exemplified by attempts to show base-rate neglect with social phenomena (Borgida & Brekke, 1981; Kassin, 1979). Ajzen and Fishbein (1975), among others, have suggested that base-rate neglect is analogous to insensitivity to consensus information in attribution theory. This linkage was reinforced by the finding subjects attend to causal base rates (Ajzen, 1977; Tversky & Kahneman, 1982b).

Fundamental Attribution Error

Central to attribution theory is the notion that actors and observers differ in their perception of whether the cause of a behavior is dispositional or situational (Heider, 1958; Quattrone, 1982). The fundamental attribution error or dispositional bias is the tendency to "attribute another person's behavior to his or her own dispositional qualities rather than to situational factors" (Fiske & Taylor, 1984, p. 73). Even when situational factors account for the behavior of an actor, observers may make an attribution to dispositional causes. Nisbett and Ross (1980) trace the fundamental attribution error to representativeness, that "actors ... may be a more representative cause of behavior than are situations, because it is, after all, the actor who does the acting" (p. 122), and to availability, that "the actor is an easily 'available' explanation of his action

because of his perceptual proximity to his action" (p. 122). In the same vein, Bar-Hillel and Fischhoff (1981) regard insensitivity to consensus or base-rate information to be a function of representativeness.

Attitude Attribution Paradigm. The strongest evidence for a dispositional bias comes from the attitude attribution paradigm (Ross, 1978). Initial evidence came from an experiment (Jones & Harris, 1967) in which subjects (observers) were told an essay writer had either been assigned to write a pro- or an anti-Castro essay (no choice condition) or had been free to choose either position (choice condition). Surprisingly, subjects attributed pro-Castro attitudes to an essay writer assigned to write a pro-Castro essay. "If the subjects fully understood the conditions under which the essay was written", Jones and Harris concluded, "their tendency to be affected by essay content in attributing an attitude to the target person would seem to reflect incomplete or distorted reasoning in the No Choice conditions" (p. 7).

Subsequent studies have found the tendency to ignore situational constraints and to attribute cause to dispositional factors to be a robust phenomenon (e.g., Jones, Worchel, Goethals, & Grumet, 1971; Snyder & Jones, 1974). Jones and Harris (1967) attributed the bias to the actor's behavior 'engulfing the field' of observers (Heider, 1958). This explanation was later modified to accommodate

the involvement of cognitive heuristics (Hicks, 1985), such as insufficient adjustment from anchors that direct observers toward actor's behavior and away from situational constraints (Miller, Schmidt, Meyer & Colella, 1984).

The extent to which the actor is perceived to be free to choose a behavior strongly influences attitude attributions. Cantor, Pittman and Jones (1982), for example, partitioned subjects' perceptions of the actor into high, moderate and low freedom. Dispositional attributions were made for high and moderate freedom actors, but for low freedom actors only for behaviors directly relevant to the essay topic. Fein, Hilton and Miller (1990) found that providing an ulterior motive for the actor's behavior (e.g., a desire to ingratiate) eliminated the bias.

Another element is the strength of the essay argument (Jones, Riggs & Quattrone, 1979; Miller, 1974; Miller & Roser, 1982; Schneider & Miller, 1975). Jones et al., (1971) concluded that no dispositional bias is found when an essay is not persuasive, and Jones et al., (1979) reported that the strength of the essay arguments becomes more relevant as the time from exposure to attitude rating increases.

Humans as Irrational Decision Makers

Nisbett and Ross (1980) prefaced their survey of the

"strategies and shortcomings of social judgment" with the following passage:

> One of philosophy's oldest paradoxes is the apparent contradiction between the great triumphs and the dramatic failures of the human mind. The same organism that routinely solves inferential problems too subtle and complex for the mightiest computers often makes errors in the simplest of judgments about everyday events. (xi)

Nisbett and Ross call these systematic errors, including representative thinking and the base-rate fallacy, "the ingredients of individual and collective human tragedy" (p. 251). A generation of psychologists and consumers of psychological research have been taught that humans are irrational decision makers, a view less likely to be held by more experienced colleagues (Beach, Christensen-Szalanski & Barnes, 1987).

Criticisms of the model of the irrational human decision maker emerged during the 1980's. One early and vocal critic was Cohen (1981), who stated:

> On the one hand, reputable investigators tell us that certain psychological discoveries have 'bleak implications for human rationality'... On the other hand, those investigators are reminded that people could not even drive automobiles unless they could assess uncertainties fairly accurately.

(p. 317)

Einhorn and Hogarth (1981) suggest that the emphasis on faulty reasoning has resulted in a distorted view of human judgment. The literature portrays humans as "intellectual cripples, limited in their capacity to think, and biased by cognitive processes that interfere with rational decision making" (Ebbesen & Konecni, 1980, p. 23).

Tversky and Kahneman (1983) respond that the focus on judgmental error has been a useful research strategy. There is precedent for such an approach in other areas of psychological inquiry (Kahneman, 1991), most notably perception, memory, and personality. However, Tversky and Kahneman are among a number of researchers in the area (e.g., Sherman & Corty, 1984) who have recognized the imbalanced state of the decision making literature. Nevertheless, the task of cataloguing errors in judgment still dominates the judgment literature, and the opinion expressed within that literature is that human decision makers are guilty of frequent acts of irrational decision making.

Two camps have emerged on the question of whether human decision making is rational or error prone (Einhorn & Hogarth, 1981; Jungermann, 1986). According to Jungermann, 'optimists' argue that many demonstrations of error in judgment are divorced from real-world contexts, fail to incorporate factors such as cognitive effort, and ignore

subjects' representation of the judgment task. 'Pessimists' counter that laboratory research reflects real-world decision processes, and that the focus on judgmental bias has provided valuable insights into the nature of human decision making.

Limitations to Normative Models

The two camps disagree on whether any normative model can serve as the standard for judging the accuracy of human decision making. An extreme view is that because all normative models are the product of human reasoning, such models cannot then serve as the standard by which the rationality of human reasoning is assessed (Cohen, 1981; Frisch, 1988). A more moderate view is that normative models are arbitrary criteria (Kruglanski & Ajzen, 1983) that may serve as a point of comparison.

Tversky and Kahneman (1983) acknowledge the difficulty in establishing a criteria for evaluating human judgment. In contrasting probability tasks to perceptual illusions, Tversky and Kahneman concede that "because individuals who have different knowledge or who hold different beliefs must be allowed to assign different probabilities to the same event, no single value can be correct for all people." (p. 293)

Nevertheless, most members of the pessimist camp (including Tversky and Kahneman) have accepted the argument that human judgment can be evaluated against normative

models. Bayes' theorem is the most widely accepted normative standard for probability judgments. Ironically, Gigerenzer and Murray (1987) suggest Reverend Bayes was more cautious about applying his theorem than psychologists have been. In applying Bayes' theorem to derive a 'correct' answer to judgment tasks, two assumptions of isomorphism are made (Gigerenzer & Murray, 1987). Concept isomorphism is the notion that both the base-rate and likelihood components of Bayes' theorem can each be "unequivocally matched with one semantic concept concerning the problem" (p. 163). Structural isomorphism is the notion "that the formal structure underlying Bayes' theorem is represented by a similar structure of the problem" (p. 163), that there are "mutually exclusive and exhaustive hypotheses, that the data has been obtained by random sampling, and that successive data are mutually independent" (p. 163).

The original version of the Cab problem violates concept isomorphism (Gigerenzer & Murray, 1987). A witness is said to make a correct identification 80 percent of the time. This statement should be interpreted as the witness is correct about 80 percent of the time regardless of the color of the cab. A subject may assume the witness is better at identifying green cabs because, for example, green cabs are easier to see at night. A more serious violation of concept isomorphism is that either 'hit and run accidents in the city' or 'hit and run accidents at night' might be a

more appropriate base-rate than number of cabs in the city. Bayes' theorem cannot provide an answer to a problem if the correct base-rate cannot be identified.

Violations of concept isomorphism can be dealt with by adding statements that clarify ambiguity or by rewording the problem. Violations of structural isomorphism are less easily addressed. One violation of structural isomorphism identified by Gigerenzer and Murray (1987) concerns random sampling, a necessary condition for the application of Bayes' theorem (see Gigerenzer, Swijtink, Porter, Daston, Beatty & Kruger, 1989). Many judgment problems appear to make implicit reference to nonrandom sampling. For example, the personality sketches in the Tom W. were not said to be selected randomly. The personality description of Jack was obviously written to portray a stereotypic engineer rather than drawn at random as subjects are told. Indeed, Gigerenzer, et al., (1989) found less bias for the Jack problem when random sampling was made apparent to subjects.

Although Bayes' theorem is the most popular normative formulation among psychologists, it is only one of a number of models of probability. Contrary to Bayes' theorem, signal detection theory views the likelihood ratio to be dependent on the base-rate, and gives an answer to the cab problem closer to the median response of subjects than the answer provided by Bayes' theorem (Birnbaum, 1983). Baconian probability theory assesses the extent elements

within a problem are specified and the relationships between these elements (Cohen, 1981). A Baconian analysis of the Jack problem would have as an "expectation that a person's hobbies, abilities and character traits are causally linked with the choice of a profession" (Barnes, 1985, p. 26). Gigerenzer (1991a) suggests a frequentist interpretation of probability, the form of probability most familiar to psychologists (e.g., Type I error), would see it to be a matter of conjecture whether any particular person is a bank teller and/or a feminist.

These concerns extend to social judgment tasks. Barnes (1985) notes how philosophers of science have been unable to agree on concepts such as cause and randomness. Nevertheless, researchers in attribution theory expect subjects to conform to their definitions of these concepts. For example, twenty years after Kelley's (1972) attribution formulation first appeared, debate continues on what is the appropriate attribution given a set of observations (e.g., Hilton & Slugoski, 1986).

Ecological Validity

One criticism often made of psychological research is that it lacks ecological validity (e.g., Campbell, 1966; Farrington, 1980; Scheidt, 1981). Toda (quoted in Barnes, 1985) concluded "man and rat are both incredibly stupid in an experimental room", and there has been "little attention [paid] to the things they do in their normal habitats." (p.

3) Some members of the optimist camp (e.g., Funder, 1987; Hogarth, 1981; Jungermann, 1986) have directed this general criticism of laboratory research toward demonstrations of faulty human judgment, suggesting such errors arise only under special circumstances. Nisbett and Ross (1980) respond that laboratory demonstrations underestimate the extent of error, and express their "impatience with those who suggest that systematic inferential failings are restricted to exotic laboratory contexts" (p. 252).

Urn-and-balls problems were discarded, in part, because of their apparent lack of relevance to real-world problem solving and their failure to involve subjects. In addition to including real-world elements in their problems (e.g., taxicabs), Kahneman and Tversky (1972) engaged in "special efforts" to maintain "the attention and motivation" of their subjects (p. 431). For each demonstration, an attempt was made to simulate a "natural classroom situation" by presenting the problems in a "quiz-like fashion" (p. 432).

Rather than problem content or administration, the ecological validity concern is whether the same or equivalent decision making processes are invoked in the laboratory as in the real-world. Hogarth (1981) argues laboratory tasks misrepresent the nature of real-world decision making. Decisions that reflect continuous processing of information may appear to be in error if evaluated in a context that demands discrete processing

(Jungermann, 1986). For example, the most available instances may be recalled to serve as anchors. If feedback is continuous, instances shown to be inadequate anchors will be discarded. Discrete laboratory tasks do not permit corrective activity of this type.

In laboratory decision making tasks, subjects are often provided with only the information the experimenter judges to be relevant (Dawes, 1988). In the real-world, other information may lead to a different judgment. Ebbesen and Konecni (1980) argue the decision making literature has concluded subjects are error prone based on tasks "decomposed into the dimensions that [are] of primary interest to the researcher" (p. 21) rather than of importance to the subjects. Furthermore, what experimenters define as error in the laboratory may reflect optimal performance in the real-world if evaluated in terms of costs and benefits (Hogarth, 1981). If judgmental accuracy is secondary to some other concern such as conservation of effort, then a judgment that reflects such a concern is not in error (Barnes, 1985).

Context

Optimists argue the decision literature has ignored the context in which the decision takes place. Defining representativeness as a "general, task-independent heuristic, implies that the context of the problem is not very important." (Gigerenzer & Murray, 1987, p. 156)

Berkeley and Humphreys (1982) have concluded decision theorists have failed to explore "small worlds", the context for the decision making task that the subject has defined. In a similar vein, Wimsatt (1986) believes decision making processes are "designed for use in certain contexts and can produce biased or worthless results if they are used in contexts that may fail to meet the conditions for which they were designed." (p. 297)

Margolis (1987) argues that many so-called errors in judgment arise from contextual ambiguity in the language or setting. For example, Margolis sees the Cab problem as reflecting a scenario effect resulting from ambiguity in language. The original version of the Cab problem may not lead subjects to infer that a green cab incorrectly identified must therefore be blue. The original version also refers to the probability a cab was identified as blue and not that the actual color of the cab was blue. The following modifications, Margolis argues, corrects this ambiguity:

> 85% of the cabs in the city are Green and 15% are Blue; and the maintenance policies, management, and so on of the two companies are similar...A bystander identified the cab as Blue. The police tested his reliability under the same circumstances as existed on the night of the accident. The tests showed that he correctly

> identified Blue cabs as blue 80% of the time, but misidentified Green cabs as blue 20% of the time. (p. 160)

Margolis (1987) also detects scenario effects within the Linda problem. Asking which is more probable, an event A or the conjunction of two events A and B, implies that an event and its conjunction are mutually exclusive. Thus, the statement 'Linda is a bank teller' might be read as 'Linda is a bank teller and not active in the feminist movement'. Margolis (1987) rewrote the Linda problem to deal with some of these contextual ambiguities and did not find the bias:

> Linda is 31 year old, bright and outspoken. As an undergraduate she majored in philosophy and was active in the environmental and civil rights movements. A personnel survey showed that of clerical workers in banks (including tellers) fewer than 1% have personality profiles that sound similar to Linda's.
>
> If you stood to win $10 if the statement you choose turns out to be true (whether or not the other statement is also true), which choice is more likely to win you the $10.... (p. 166)

Subjects' Interpretations

An understanding of how decisions are made requires more than merely examining the judgment given. It requires knowing "(a) how subjects structure the problem and (b) how

they appraise the different components of the inferential model they are using" (Fischhoff & Beyth-Marom, 1983, p. 251). An examination of how the task is interpreted is required prior to concluding that judgment is biased (Cohen, 1982; Sherman & Corty, 1984).

One objective for such an examination should be to determine if subjects' perceive the judgment task in the way the experimenter intended. An "incorrect" response to a problem could be the product of either faulty information processing or choice of problem solving strategy (Pitz, 1980). It can only be the latter if the experimenter and subjects understand the task in the same way. If it somehow becomes apparent that the experimental task is being perceived by subjects differently from what was intended by the experimenter, however, the conclusion often drawn is subjects have misrepresented the task rather than having devised a strategy "best suited to solving the problem as he or she perceives it" (Pitz, 1980, p. 187).

Karlsson (1989) traces this confusion between subject and experimenter to a notion of William James, 'the psychologist's fallacy'. "The psychologist mistakes his/her knowledge about a situation ... for the subject's experience of it" (Karlsson, 1989, p. 54). Faulty reasoning may reflect "normatively plausible responses to a question different from what the experimenters intended" (Margolis, 1987, p. 141). In his evaluation of the social judgment

literature, Funder (1987) views any experiment in human judgment as an interpersonal "interaction among the subject, the experimenter, and even the institution ... (cf. Orne, 1962)" (pp. 79-80). And Fischhoff (1990) speaks of "opening this can of subjects" and finding out "what subjects believe our games to be all about?" (p. 341).

The paradigm employed to test for the involvement of heuristics in decision making contributes to uncertainty in interpreting the behavior of subjects. The cooperativeness principle (Grice, 1975; Turnbull & Slugoski, 1988) states that the listener in a conversation should expect the communicator to be truthful and to provide only relevant information. Kahneman and Tversky (1982) suggest this principle may impact on subjects' interpretation of judgment tasks involving irrelevant information (e.g., a personality description) "because the presentation of irrelevant information violates rules of conversation, [and] subjects are likely to seek relevance in any experimental message. (p. 502) The cooperativeness principle suggests that subjects will interpret these descriptions as having been given for some purpose, most likely to assist them in judging the occupation of the target individual (Adler, 1984).

Evaluation of Subjects' Interpretations. Researchers have called for studies to evaluate subjects' interpretation of base-rate judgment problems (e.g., Berkeley & Humphreys,

1982; Cohen, 1982; Jungermann, 1986). There have been comparatively few studies that have undertaken this task. Lyon and Slovic (1976) asked subjects to check which of three interpretations of the Cab problem they held. Of the twenty-seven subjects in the study, fourteen chose an interpretation that endorsed likelihood information and rejected base-rate information as relevant to solving the problem. Only four subjects chose a base-rate interpretation, and only nine selected an interpretation that combined base-rate and likelihood information. Thus, only approximately half of the subjects in the Lyon and Slovic study chose an interpretation of the Cab problem consistent with that held by researchers.

Kahneman and Tversky (1982) investigated research strategies used in the Linda problem. Subjects were presented with two interpretations of the problem. One interpretation was based on probability theory, the other, the interpretation most commonly selected, focused on the description given of Linda.

Barnes (1985) differentiated aleatory logic, which emphasizes characteristics of a set of objects, from epistemic logic, which focuses on the characteristics of each object of a set. She argued that although the research literature assumes aleatory logic will be applied to judgment tasks, subjects may interpret these problems as requiring epistemic logic. To test this, subjects were

asked to sort problems on cards into piles according to the method that might be used to solve the problem. Subjects' categorization of the problems did not correspond to the interpretations of the problems suggested by the literature (e.g., base-rate). Analysis of think-aloud protocols showed the Linda, Jack, and Tom W. problems demanded epistemic rather than aleatory logic.

A number of more recent studies have replicated the finding that subjects hold a variety of different interpretations of judgment problems rather than the single interpretation dictated by researchers and the normative model (e.g., Gigerenzer et al., 1988; Konold, 1989; Lindeman, Van den Brink, & Hoogstraten, 1988; Scholz, 1987; Yates & Carlson, 1986). Less direct evidence comes from the striking variability in subjects' responses to these problems. Large variability within- and between-subjects, and across problems have been observed for both probabilistic (e.g., Barnes, 1985; Evans & Pollard, 1982; Gigerenzer & Murray, 1987) and social (e.g., Jones & Harris, 1967) decision tasks.

Two-Experiment Problem

Kahneman and Tversky (1982) warned of miscommunication where the subject and experimenter hold different representations of the same judgment task. Orne (1973) labelled this a "two-experiment problem". Any laboratory study can be viewed as two distinct experiments -- the one

"conceived by the experimenter", the other "experienced by the subject" (p. 161). The greater the similarity between these conceptions of the experiment, the stronger the conclusions that can be derived from the research.

Subjects' perceptions of an experiment arise from two sources: "(a) those which are traditionally defined as experimental variables, and (b) the perceived demand characteristics of the experimental situation" (Orne, 1962, p. 779). The two-experiment problem emerges when the demand characteristics suggest an interpretation of the situation inconsistent with the experimenter's conceptualization. Quasi-control techniques (Orne, 1969), including post-experimental questionnaires, non-experiments, and simulating subjects, were proposed as ways to evaluate the nature of the experimental situation.

This view did not originate with Orne. In the Gestalt tradition, a precursor to modern cognitive psychology, it was recognized that sometimes subjects' view of a task may not have "much in common with the problem with which the investigator started" (Asch, 1952, p. 405). This discrepancy between experimenter and subjects was referred to in the concept attainment literature as a "'two-man game' feature of most experimental research on the thought processes" (Bruner, Goodnow & Austin, 1956, p. 58). Kahneman and Tversky (1982) echoed these concerns in the context of judgmental biases. According to their view,

judgmental errors "could also arise from the subject's misunderstanding of the question or from the investigator's misinterpretation of the answer" (p. 493). Kahneman and Tversky cautioned against assuming "that all the information given to the subject is included in the experimental message" (p. 502). Subjects may be raising questions that the "experimenter never thought of asking" (p. 502):

> Is there a correct answer to this question? Does the experimenter expect me to find it? Is an obvious answer at all likely to be correct? Does the question provide any hints about the expected answer? What determined the selection of the information that I was given? Is one of it irrelevant and included just to mislead, or is it all relevant? (pp. 502-503)

The demands of the experimental situation are the source of this miscommunication (Kahneman & Tversky, 1982).

There have been few recent systematic investigations into demand characteristics within experimental psychology (Greenberg & Folger, 1988). The absence of research can be attributed to continued domination by logical positivism (Harre & Secord, 1972), "resolution" of the crisis of confidence in most fields of psychology (Adair, 1991), and arguments against the demand characteristics perspective (Berkowitz & Donnerstein, 1982; Berkowitz & Troccoli, 1986).

Critics of demand characteristics (e.g., Berkowitz &

Troccoli, 1986) argue that subjects do not engage in an "active intellectual search for the study's true purpose" (p. 341) and are not "highly motivated to ensure the study's success ... intent on obtaining evidence regarding the investigation's true purpose ... [nor] eager to confirm this hypothesis, supposedly in the interest of science" (p. 343). Proponents of demand characteristics (e.g., Adair, 1982) suggest critics have focused on a narrow and outdated conceptualization of the demand thesis. Rather than subjects actively seeking to ascertain the researcher's hypothesis, demand characteristics more correctly refer to subjects' search for the meaning of the experiment. "This meaning may be the same as the experimenter's hypothesis or it may be different, in which case the two-experiment problem arises." (Adair, 1982, p. 1406)

Demand characteristics have been only indirectly linked to judgmental bias (c.f., Adair, 1991; Kahneman & Tversky, 1982). Maxims of communication such as the cooperativeness principle (Grice, 1975) have increasingly been considered within the bias literature (e.g., Kahneman & Tversky, 1982; Krosnick et al., 1990), although rarely investigated (Gigerenzer, 1991b). Violations of these linguistic principles are demand characteristics that cause subjects to interpret the research task in a different way than intended by the experimenter (the two-experiment problem).

An advantage to interpreting laboratory investigations

of judgment bias from the two-experiment perspective is that it highlights limitations of previous research and suggests a framework and methodology for inquiry. One such limitation to previous research has been the preoccupation with "evidential variables" (Tversky & Kahneman, 1982a), the variables experimenters deem to be of consequence, specifically the base-rate and likelihood information. What have been ignored are "procedural" variables, such as the context of the problem as perceived by the subjects. A related limitation to previous research has been the strong inclination to examine only binary or numerical responses as dependent variables. The result is preoccupation with the question of whether subjects deviate from normative solutions (Gigerenzer, 1991a).

One intermediate research strategy that seeks to ascertain why subjects deviate from normative solutions is content analysis of verbal protocols. This approach has gained favour with both decision researchers (Montgomery & Svenson, 1989) and demand characteristic proponents (Adair, 1991). However, content analysis rarely attempts to ascertain the meaning subjects attribute to problems (Svenson, 1989), and often results in complex and uninformative data (Edwards, 1982).

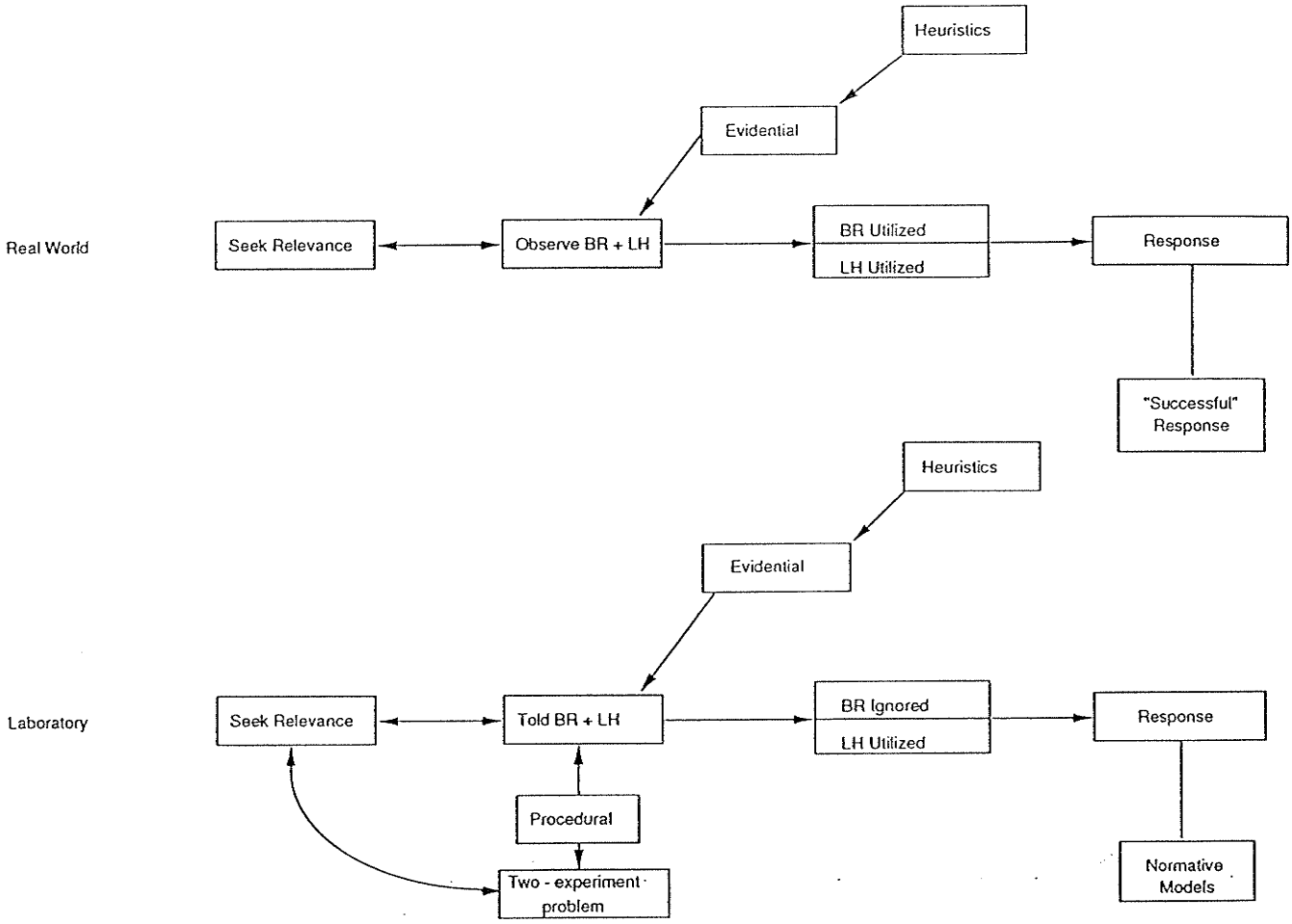
A two-experiment analysis of the bias literature suggests a different research strategy. The investigation should focus on what Tversky and Kahneman (1982a) referred

to as procedural variables, including "the properties of the design, the task, and the display" (p. 154), but also the instructions to subjects as "the greatest single source of artifact in laboratory experiments" (Adair, 1991, p. 447). Properties of the experimental procedure and experimental instructions convey to subjects the meaning and purpose of their participation in the research. One should ascertain what subjects believe the objective of the research to be, and what elements of the task they perceive to be relevant. Dependent variables should include not only subjects' responses to the judgment problems, but also phenomenological measures to assess their perceptions of the experimental task and context.

Statement of the Problem

A framework for representing the base-rate judgment process is presented in Figure 1. This model summarizes the judgment process, examines how miscommunication may arise, and contrasts judgment in laboratory and real-world settings. Decision making theorists (e.g., Kahneman and Tversky, 1982), philosophers of cognitive science (e.g., Sperber & Wilson, 1991), linguists (e.g., Roberts, 1991), and research methodologists (e.g., Adair, 1991), have suggested problem solvers seek relevance or meaning for their participation in a task and the information they are provided. This tendency would apply to evaluation of base-rate (BR) and likelihood (LH) information.

Figure 1: A Conceptual Model of Base-rate Judgments in the Real-World and the Laboratory



Significant differences between real-world and laboratory decision making are evident at this stage of the model. Problem solvers in the real-world are forced to select, interpret, and interact with information from the problem-solving environment (Dawes, 1988). In contrast, laboratory subjects are provided the information the experimenter judges to be important in written or spoken form. These differences may result in a distortion of the decision making process (Kahneman & Tversky, 1982).

Heuristics such as representativeness are seen to impact on how base-rate and likelihood information are integrated. The interaction of heuristics and base-rate/likelihood information encompass evidential variables (Tversky & Kahneman, 1982a). Critics (e.g., Gigerenzer, 1991b) argue the singular focus on evidential variables has severely limited understanding of the decision making process. In contrast, procedural variables include "the nature of the problem, the characteristics of the design, the sophistication of the respondents, and **the presence of suggestive clues or other demand characteristics**" (Tversky & Kahneman, 1982a, p. 154, emphasis added). Procedural variables include demand characteristics (Orne, 1962), but also conversational maxims such as the cooperativeness principle (Grice, 1975), and the two-experiment problem (Orne, 1973). These variables affect subjects' general view of the experimental task, as well as specific

interpretations of base-rate/likelihood information, and interact with subjects' general tendency to seek relevance from their participation in the task. It follows that the information perceived to be relevant by subjects according to their understanding of the experimental task has greatest impact on the decision outcome.

The decision outcome within the model highlights two other differences between real-world and laboratory contexts. The first is that robust laboratory demonstrations of base-rate underutilization in the laboratory (Borgida & Brekke, 1981) conflict with the significant impact baserates sometimes have in real-world decision making (Gigerenzer, 1991b; Nelson, Biernat & Manis, 1990). Individuals are often judged on the basis of characteristics presumed to be shared by members of their social group. These stereotypes are baserates that are "potent and resilient" to change (Nelson et al., 1990, p. 664). The second is that the criteria for a "correct" response in the laboratory is narrow and arbitrary compared to the criteria in the real-world. In the laboratory, normative formulations such as probability theory dictate the "correct", albeit equivocal solution to a problem. In the real-world, many factors must be evaluated in order to determine what is a "successful" outcome (Gigerenzer, 1991b).

Before condemning subjects' judgment capabilities in

the laboratory compared to that in the real-world, it is essential to better understand how procedural variables impact on subjects' perceptions of the experimental situation and the decision process. Only a handful of studies (e.g., Barnes, 1985; Gigerenzer et al., 1988; Scholz, 1987; Schwarz, Strack, Hilton & Naderer, 1991) have systematically examined these situational (e.g., nature of the problem) and individual difference variables (e.g., sophistication of the respondents; Tversky & Kahneman, 1982b). To conclude that judgment is systematically biased, it must be shown first that subjects interpret these problems in a manner consistent with the interpretation intended by the experimenter, and then that they deviate from the normative model. Applying a two-experiment analysis to classic demonstrations of the base-rate fallacy provides insight into the sources of these biases (Bar-Hillel, 1990).

A series of experiments were designed to systematically address these issues. Five problems were selected for study: the Cab, Jack, Tom W., and Linda problems, and an attitude attribution problem ("Bret"). These problems directly relate to application of Bayes' theorem to human decision making because deviations from normative solutions are seen to result from overreliance on likelihood information and underreliance on base-rate information, the two components of the Bayes' formula. As well, subjects'

apparent exclusion of base-rate information from consideration appears in both the probabilistic and social domains. The base-rate problems selected for this investigation sample from both domains. Utilization of base-rate information is also a significant issue in real-world contexts such as jury deliberations (e.g., Hastie, Penrod, & Pennington, 1983) and clinical judgments (e.g., Meehl, 1957).

The first experiment was somewhat exploratory --- examining a number of basic questions about subjects' interpretation and response to typical judgment problems. For example, subjects' post-experimental perceptions of the experimental task were examined to see if they related to their responses to the problems. A post-experimental questionnaire (PEQ) asked subjects to express their general understanding of the experimental task, and their specific appraisal of each problem. These were compared to the interpretation demanded by the literature and dictated by the normative model.

If the interpretation of the experimental task held by subjects differed from that intended by experimenters, then the next task would be to identify the factors that might lead to development of these alternative interpretations. Because previous investigations have found substantial variability in responses to the same problem, individual difference variables possibly mediate different

interpretations of the same problem. Measures of personality, cognitive functioning, experimental attitudes, and statistical sophistication were assessed as possible mediators of such differences in Experiment 1.

Past research has focused on the role of evidential variables, specifically the manipulation of base-rate and likelihood values, in order to demonstrate the effects of judgmental heuristics. This approach conflicts with a model that views procedural variables and subjects' interpretations of the task as significant mediators. Because subjects appear to manipulate evidential variables inappropriately (e.g., the base-rate; Tversky & Kahneman, 1982a), and to be less affected by procedural variables (e.g., the context of the problem; c.f. Gigerenzer & Murray, 1987), responses to one base-rate problem are generally assumed <u>not</u> to affect performance on subsequent problems. Because heuristics are regarded as task-independent (c.f. Gigerenzer & Murray, 1987), then subjects who deviate from the normative response for one problem should deviate from the normative response for other problems. On the other hand, subjects may respond differently to different problems, or subjects may be affected by order of presentation of the problems. Either outcome would imply that unique characteristics of each problem lead subjects to deviate from normative solutions, and provide support for the importance of procedural variables in Experiment 1.

In order to solve a base-rate judgment problem, the appropriate base-rate and likelihood information must be correctly identified. Even if the experimental task was interpreted in a manner consistent with the normative model, an incorrect response may be given because the subject is unable to identify the appropriate base-rate or likelihood information. Subjects should have little difficulty identifying the appropriate likelihood information (e.g., the personality description) because it is highly salient and vivid (Bar-Hillel, 1980), and because experimental demands direct subjects toward such information. However, base-rate information is often presented in a sentence at the beginning of a problem (e.g., the Jack problem), or must be derived from world knowledge (e.g., the Tom problem). The second experiment sought to ascertain whether subjects correctly identify the appropriate base-rate and likelihood information for each problem, and whether being directed toward the base-rate information by such a task resulted in less error.

Alternative versions of each problem that have been shown to produce smaller deviations from the normative response are thought to influence primarily evidential variables (e.g., making the base-rate causal). On the other hand, these problems may suggest to subjects interpretations of the problems more consistent with those held by the experimenter. The third experiment tested the differences

in performance and views of alternative and standard problems with phenomenological inquiries narrowed to provide subjects with an opportunity to express their views of each problem immediately after offering a solution. The hypothesis tested was that alternative versions of each problem may reduce deviations from normative solutions by affecting subjects' perceptions of the problem (e.g., suggesting an interpretation more consistent with that held by the experimenter) rather than by affecting the evidential variables (e.g., the base-rate).

Subjects may be guided to adopt an interpretation of the experimental task consistent with that held by the experimenter if demands to do so were made explicit. The fourth experiment was designed to manipulate a procedural variable, the problem frame or the demands of the experimental task, without altering the base-rate and likelihood information. It was hypothesized that these manipulations would result in smaller deviations from the normative response.

In the real-world, many significant decisions are made by groups of individuals engaged in social interactions. Group discussion offers a forum for expression and validation of different interpretations of a problem. It directs attention to the task at hand, permits the critical evaluation of responses, forces defence of a position, and demands consideration of other responses. If subjects are

failing to provide the normative response to a base-rate problem because of a lack of motivation or cognitive effort (e.g., Langer, 1989), then group discussion might provide an incentive for subjects to apply themselves to the experimental task. Alternatively, group discussion may provide an opportunity for subjects who interpret the task in a manner different from that intended by the experimenter to be exposed to the normative interpretation for the task. This approach might reduce what Kahneman and Tversky (1982) referred to as errors of application, the knowledge of the appropriate rule for solving a problem but failing to recognize the rule applies to the particular circumstance. The fifth experiment considered the influence of group discussion on subjects' judgments and perceptions.

EXPERIMENT 1

The defining characteristic of a two-experiment problem is the miscommunication between experimenter and subject (Orne, 1973). An initial first step in understanding the nature of the interaction between subjects and experimenter is to evaluate how subjects interpret the experimental task. To begin the phenomenological analysis of subjects' perceptions, a sophisticated post-experimental questionnaire was used in Experiment 1 to assess whether subjects whose judgments deviated from the normative (e.g., Bayesian) model differed in their interpretation of the task from subjects whose responses were 'correct', that is from the interpretation intended by the experimenter.

A normative interpretation of the task demands that there be only one correct solution to a problem, and that the solution be obtained through formal analysis (e.g., probability theory). Subjects were asked whether a problem had more than one correct response, and whether an intuitive or formal analysis was required. It was predicted that subjects who perceived a problem in a manner consistent with the normative interpretation would deviate less from the "correct" response to the problem.

The "correct" response to each problem also demands baserates be seen as relevant, and likelihood information be ignored. However, baserates may be neglected if seen as irrelevant to solving the problem (Bar-Hillel, 1990; Schwarz

et al., 1991). In this experiment, subjects were asked to evaluate base-rate and likelihood elements of each problem for their relevance to solving the problem. Subjects who perceived the base-rate information to be relevant were hypothesized to deviate significantly less from the normative solution for each of the five problems.

Conceptualizing heuristics such as representativeness as context-independent (Tversky & Kahneman, 1974), implies that aspects of the experimental setting, such as the presentation of multiple problems to the same subjects, are of little consequence and probably do not affect subjects' perceptions of the experimental task. However, larger deviations from normative solutions have been found for within- over between-subject designs (Tversky & Kahneman, 1982a; c.f., Fischhoff, 1977), and concern has been expressed that within-subject designs are more susceptible to experimental demands (Kahneman & Tversky, 1982; Krosnick & Sedikides, 1990). In this vein, Gigerenzer et al., (1988) found a complex effect for order of placement for the Dick version of the engineer/lawyer problem, and both Hinsz et al., (1988), and Scholz (1987) reported a non-significant trend for problems presented later in a set to deviate less from the normative standard. Although not explicitly provided with performance feedback, a subject may 'catch on' to the demands of the experimental task (e.g., to ignore the individuating information; Kahneman & Tversky, 1982). Thus,

among the set of five problems administered in Experiment 1, deviations from the normative solution were predicted to become progressively smaller the later a problem appeared within the set. However, none of the studies that have presented more than one base-rate problem to the same subjects (e.g., Barnes, 1985) have explicitly examined the inter-relationship between performance on different problems. Because all five problems in Experiment 1 have been linked to base-rate neglect and related phenomena, it was hypothesized that deviations from the normative solutions for the five problems should intercorrelate.

The emphasis on manipulating base-rate and likelihood information ignores the substantial variability in subjects' responses to the same problem (Furnham, Jaspars & Fincham, 1983), and therefore the individual difference variables that may lead some subjects to form a different interpretation of the task than the interpretation held by other subjects and intended by the experimenter. One personality variable related to social attributions is self-monitoring (Snyder, 1974). Krosnick and Sedikides (1990) found high self-monitors, individuals who monitor social cues in order to project a positive self-image, were more sensitive to 'complimentary' base-rate information that predicted the subject would offer assistance in a helping situation. On the other hand, low self-monitors, individuals who try to present their 'true' selves, were

more sensitive to 'threatening' base-rate information that predicted the subject would not offer assistance. Because the base-rate information in this study is neither complimentary or threatening to subjects, no specific prediction of the relationship between self-monitoring and absolute deviation scores was made.

Another individual difference variable that may affect subjects' perceptions of the problems is attributional complexity (Fletcher, Danilovics, Fernandez, Peterson and Reeder, 1986) or the level of sophistication of the causes made for behavior. Fletcher et al., predicted that attributionally-complex individuals should be less likely to produce errors in social judgment, and Devine (1989) found attributionally-complex subjects produced a significantly smaller dispositional bias for a measure of subject confidence. In contrast, Fletcher, Reeder and Bull (1990) found attributionally-complex subjects to be somewhat more likely to show a dispositional bias on the traditional measure of attitude extremity. It was hypothesized that because attributionally-complex subjects may search likelihood information (e.g., a personality description) for clues as to the dispositional cause for a target's behavior, attributional complexity would correlate positively with deviation scores for the Jack, Linda and Tom problems. On the other hand, because attributionally complex subjects may also be drawn to base-rate information that suggests a

situational explanation for a target's behavior (e.g., assigned to write a counterattitudinal essay), attributional complexity would correlate negatively with deviation scores for the Bret problem.

Related to attributional complexity is the idea that individuals differ in their level of engagement in cognitive activity or need for cognition (Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1984). Although Allwood and Bjorhag (1990) did not find a relationship between the Need for Cognition scale and confidence judgments, items such as 'I prefer my life to be filled with puzzles that I must solve' and 'I would prefer complex to simple problems' implies cognitive complexity should relate to performance on judgment problems. As with the attributional complexity contrast, greater deviation from the normative solution was predicted for subjects high in cognitive complexity for "personality" problems (e.g., Jack, Linda, and Tom), but lesser deviation for more "task-oriented" problems (e.g., Bret).

The final individual difference variable considered was subjects' attitudes toward psychological research (Psychological Research Survey; Adair & Fenton, 1971). Subjects with more positive attitudes toward psychological research should be more susceptible to experimental demands to produce behavior consistent with their understanding of the hypothesis of the study. If within base-rate problems,

there are experimental demands to utilize the likelihood information and to ignore the base-rate information, and subjects correctly interpret these cues, then positive perceptions of psychologists and psychological research should be associated with deviations from normative solutions for the five problems.

Method

Subjects. One-hundred and twenty first-year introductory psychology students from the subject pool at the University of Manitoba participated in Experiment 1. Subjects received credit for their research participation.

Materials and procedure. Subjects were gathered in groups of about thirty to complete the following measures in order: (1) the Psychological Research Survey, (2) the Attributional Complexity scale, (3) the Need for Cognition scale, and (4) the Self-Monitoring Scale. All four scales appeared consecutively in an experimental booklet (Appendix A). Responses were recorded by subjects on I.B.M. answer sheets. Half of the subjects completed the personality measures before attempting the judgment problems.

A separate experimental booklet (Appendix B) contained the five judgment problems. Each of the 120 unique orders of the five problems was used so each problem appeared in the same position (e.g., the first problem attempted) in the same number of booklets. Subjects were explicitly instructed on the first page of the experimental booklet not

to return to an earlier problem nor to flip ahead through the booklet. The experimental booklet contained standard versions of the Cab, Jack, Tom, and Linda problems described in detail in the Introduction. The response format for the Linda and Tom problems was modified to permit subjects to respond on a continuous scale (e.g., Fischhoff & Bar-Hillel, 1984b). For example, responses to the Linda problem were placed on a 7-point scale, between the anchors 'Bank Teller' and 'Bank Teller and Feminist'.

A fifth problem presented in the experimental booklet, the Bret problem, was constructed especially for this study. It required subjects to determine the attitude of a student named Bret ostensibly assigned to write an essay <u>against</u> funding for research into Acquired Immune Deficiency Syndrome (AIDS). AIDS has been used in a previous attitude attribution study (Devine, 1989), and pilot studies showed the Bret problem reliably induces a dispositional bias.

Each problem was presented on a separate page of the booklet, and each subject was requested to rate their confidence in their judgment on seven-point scales from 1 'Very Unsure' to 7 'Very Sure'. The experimental booklet concluded with a post-experimental questionnaire. Subjects were first asked general questions to ascertain their views of the experimental task. These included the purpose of the experiment, the reason for being asked to solve the problems, and the quality or skill being assessed. Subjects

were also asked if any of the problems did not have a correct response and if there was a trick to solving the problems.

These general questions were followed by specific questions responded to on appropriately anchored 7-point scales and accompanied by a reproduction of each problem to aid recall. Subjects were asked whether each problem had a single correct response or more than one correct response, and whether each problem required an intuitive solution or a more formal analysis. Subjects were also asked to rate elements from each problem for their relevance to arriving at a solution. These elements were created by partitioning each problem into one or two sentence statements that reflected either base-rate or likelihood elements.

Results

Responses to the Judgment Problems. Before examining results pertaining to specific hypotheses, it must be shown that subjects in Experiment 1 replicated previous studies by producing responses that deviated significantly from normative solutions. Summary statistics were calculated for each of the five problems (Table 1). The first, third, and fourth rows of Table 1 present, respectively, the mean, median and modal responses to each problem.

Across all five problems, responses were consistent with those from previous studies. For each of the problems,

Table 1

Summary Statistics for Responses to the Five Judgment Problems, Experiment 1

	Cab	Jack	Linda	Tom	Bret
M	58.91	65.56	5.45	5.83	2.66
S	28.93	28.83	1.52	1.19	1.55
Median	75	75	6	6	2
Mode	80	30	6	6	2
Skewness	-.82	-.41	-1.14	-1.63	.83
M_d	32.67	37.09	4.45	4.83	4.34
S_d	11.00	26.82	1.51	1.19	1.55

Note. *N* = approximately 120. *M* = mean response, *S* = standard deviation for the responses, M_d = mean absolute deviation score, S_d = standard deviation for the absolute deviation scores

however, the mean, median, and modal responses were not the same. Kolmogorov-Smirnov $\underline{D}$ statistics ($\underline{p} < .01$) confirmed responses to each of the five problems were non-normally distributed. Non-normal distributions can affect statistical analyses (Levine & Dunlap, 1982; Tabachnick and Fidell, 1989). Tabachnick and Fidell (1989) recommend reflecting the responses to negatively skewed distributions (done for all but the Bret problem which was positively skewed), followed by square root transformations to reduce skewness. However, this approach did not alter the outcome of the statistical tests to follow.

Table 1 also provides means and standard deviations for the absolute difference of each subjects' response from the normative solution ($\underline{M}_d$ and $\underline{S}_d$; Scholz, 1987). The absolute difference was calculated so that positive and negative signed deviations from the normative solution would not cancel each other out. The normative solution for the Cab problem (the probability the cab was blue) was 41 percent. The probability that Jack was one of the 30 engineers in the sample of 100 was 30 or the base-rate. A rating of 1 on the seven-point scale was the normative solution to both the Linda problem (a bank teller) and the Tom problem (a graduate student in education). A rating of 7 on the seven point-scale (the author of the essay is strongly <u>for</u> funding for AIDS research) was the normative solution to the Bret problem. The larger the value of the absolute deviation

score, the greater the difference between a subject's response and the normative solution. Thus, consistent with administration of these problems in previous studies, subjects deviated significantly from the normative solutions for all five judgment problems.

Nature of the Solutions. For each problem, subjects were asked whether a single correct response or more than one correct response was demanded, and whether an intuitive judgment or a formal analysis was more appropriate. A normative interpretation corresponds to a formal analysis of the problem (e.g., probability theory) rather than an intuitive analysis (e.g., representativeness), and a single correct response (e.g., the numerical value derived from probability theory) rather than multiple correct responses. It was predicted that reporting a normative interpretation for the problems would be associated with smaller deviations from the normative solution.

To permit the test of the hypothesis that subjects who endorse a normative interpretation for a problem, specifically a formal analysis with a single correct solution, median splits on the number of correct responses and the form of analysis items were used to classify subjects into one of four cells in a factorial design[6] for

[6]An alternative to a median split is to select some arbitrary value such as the mid-point of the seven-point scale. The advantage of the median split procedure is that it results in subjects being distributed somewhat equally across the four cells.

each of the five problems. Deviations from the normative solution were associated with an intuitive analysis for the Jack (F (1,116) = 20.88, MS_e = 572.09, p < .0001), Linda (F (1,114) = 3.46, MS_e = 2.27, p < .07), and Tom problems (F (1,115) = 6.56, MS_e = 1.35, p <.02), and selection of a single correct response for the Bret problem (F (1,115) = 10.66, MS_e = 2.24, p <.001). Cell means are presented in Table 2.

Within-subjects ANOVAs[7] revealed significant differences (p < .0001) between the five problems for both the number of correct responses (F (4,456) = 10.11, MS_e = 2.86) and the form of analysis items (F (4,460) = 38.07, MS_e = 3.77). For number of correct responses item, Tukey post-hoc tests (p < .05) revealed the Cab problem was associated with a single correct response (M = 2.8 on the seven-point scale) more so than either the Jack (M = 3.39), Bret (M = 3.59), Tom (M = 3.39), or Linda problems (M = 4.13). The Jack problem was also seen to require a single correct response more so than the Linda problem. For form of analysis item, the Cab problem (M = 5.54 on the seven-point scale) was seen as more consistent with a formal analysis than either the Jack (M = 4.13), Bret (M = 3.22), Tom (M = 2.93) or Linda problems (M = 2.88). The Jack

---

[7]Gender was included as a between-subject factor in this analysis. Neither main effects for gender nor the interaction of gender and problem were significant.

Table 2

Summary of Mean Absolute Deviation Scores Following Median Split for the Nature of the Solution Demanded Questions, Experiment 1

Problem	Sin	Mult	Int	Form
Cab	33.55	31.57	33.33	31.76
Jack	38.38	35.20	46.89[c]	24.89[d]
Linda	4.52	4.37	4.69	4.15
Tom	4.97	4.60	5.10[c]	4.56[d]
Bret	4.75[a]	3.85[b]	4.30	4.37

Note. Sin = single correct response, Mult = multiple correct responses, Int = intuitive analysis, Form = formal analysis. Pairs of row means with different superscripts significantly different from each other, Tukey HSD post-hoc test ($p < .05$). *ns* ranged from 45 to 75.

problem was also seen to more strongly demand a formal analysis compared to the Linda, Tom, and Bret problems.

Relevance of Problem Statements. In order to assess the importance subjects placed on the base-rate and likelihood information conveyed within each of the five problems, each problem was partitioned into base-rate or likelihood statements of one to two sentences. The Linda problem was excluded from this analysis because the base-rate component is not explicitly specified (that is, there cannot be more bank tellers and feminists than bank tellers). Subjects were asked to rate how relevant each statement was for arriving at a solution to the respective problem, and these relevance ratings were then correlated with the absolute deviation scores to determine whether subjects who endorsed the base-rate information as more relevant, and the likelihood information as less relevant, deviated less from the normative solution. Within-subject ANOVAs revealed significant differences between statements on relevance ratings for the Cab (F (1,117) = 7.78, MS_e = 3.06, $p < .01$), Jack (F (3,351) = 18.72, MS_e = 3.38, $p <$ .0001), Tom (F (5,585 = 33.89, MS_e = 2.32, p .0001), and Bret problems (F (3,354) = 5.79, MS_e = 2.53, $p <.001$).

For the Cab problem (Table 3), the base-rate statement of the number of cabs in the city was judged to be less relevant than the likelihood statement referring to the accuracy of the witness. Neither relevance ratings for the

Table 3

Relevance Rating of Statements from the Cab Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 1

Statement	M	r
85% of the cabs in the city are Green and 15% are Blue [BR].	5.24[a]	-.16
The witness made correct identifications in 80% of the cases and erred in 20% of the cases [LH].	5.88[b]	.10

Note. [BR] = Base-rate element, [LH] = Likelihood element. Mean ratings with different superscripts significantly differed according to Tukey HSD post-hoc tests ($p < .05$). * $p < .03$.

likelihood or base-rate statements were correlated with deviations from the normative response.

For the Jack problem (Table 4), statements that referred to Jack's lack of interest in political and social issues, and to Jack's hobbies, were judged most relevant. The base-rate of 30 engineers and 70 lawyers was seen as less relevant. Endorsement of the base-rate as relevant was negatively correlated, endorsement of likelihood statements were positively correlated, with deviations from the normative solution.

Statements about Tom rated as most relevant referred to aspects of Tom's personality, character, and behavior (Table 5). The base-rate statement that the description of Tom W. was based on projective tests was seen as least relevant. Again, likelihood statements were seen as most relevant, and were associated with deviations from the normative solution.

The likelihood statement by Bret that AIDS is comparatively rare was seen as less relevant than all other statements (Table 6). The 'base-rate' statement that Bret was assigned to write a brief essay was not seen as more nor less relevant than two of the likelihood statements, and was not significantly correlated with absolute deviation scores.

Thus, across the four problems, subjects either saw the base-rate statements to be of equal or lesser relevance than the likelihood statements. Perceiving likelihood statements as relevant was associated with deviations from the

Table 4

Relevance Rating of Statements from the Jack Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 1

Statement	M	r
30 engineers and 70 lawyers, all successful [BR].	3.92[b]	-.50**
He is generally conservative, careful and ambitious [LH1].	3.89[b]	.14
He shows no interest in political and social issues [LH2].	5.19[a]	.56**
He spends most of his free time on his many hobbies: home carpentry, sailing, and mathematical puzzles [LH3].	5.12[a]	.51**

Note. [BR] = Base-rate element, [LH1] = Likelihood element 1, and so on. Mean ratings with different superscripts significantly differed according to Tukey HSD post-hoc tests ($p < .05$). * $p < .005$. ** $p < .0001$.

Table 5

Relevance Rating of Statements from the Tom Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 1

Statement	M	r
The following is a description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school [BR].	3.35[e]	-.22*
Tom W. is of high intelligence, although lacking in true creativity [LH1].	4.92[bc]	.19*
He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place [LH2].	5.30[ab]	.21*
His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and flashes of imagination [LH3].	4.64[cd]	.12
He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others [LH4].	5.64[a]	.36**
Self-centered, he nonetheless has a deep moral sense [LH5].	4.20[d]	.01

Note. [BR] = Base-rate element, [LH1] = Likelihood element 1, and so on. Mean ratings with different superscripts

significantly different according to Tukey HSD post-hoc tests ($p < .05$). * $p < .05$. ** $p < .0001$.

Table 6

Relevance Rating of Statements from the Bret Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 1

Statement	M	r
Bret was assigned to write a brief essay arguing for less funding for research into AIDS [BR].	5.21[a]	.01
Bret wrote that AIDS is comparatively rare and did not warrant the coverage in the media it has received [LH1].	4.60[b]	.15
Bret wrote that an enormous amount of energy and costly research is directed toward the investigation of AIDS at the expense of more important life-threatening diseases [LH2].	5.38[a]	.19*
Bret wrote that if the same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit [LH3].	5.29[a]	.15

Note. [BR] = Base-rate element, [LH1] = Likelihood element 1, and so on. Mean ratings with different superscripts significantly differed according to Tukey HSD post-hoc tests ($p < .05$). * $p < .05$.

normative or correct responses, while perceiving base-rate statements as relevant was associated with giving the normatively correct responses to the problems.

Perceptions of the Experiment. Subjects were asked to indicate what they thought the purpose of the experiment was, why they were asked to solve the judgment problems, and which of their qualities or skills were being assessed. Most subjects made reference to judgmental ability, person perception, and confidence in their judgments. Over half of the subjects (58.6%) indicated that problems might not have had correct answers, and an additional 9.5 percent responded 'maybe'. Most subjects (60.0%) did not believe there was a trick to answering the problems. These responses were not found to be related to absolute deviation scores or confidence ratings. Only four subjects reported having seen at least one of the five problems prior to the experiment.

Correlations between the Problems. Contrary to the expectation that responses to the five problems would be correlated, the only significant correlation was between the Jack and Linda problems (r (59) = .23, p < .02). Calculation of Spearman's correlation coefficients to assess the degree of relationship between the problems regardless of the nature of the relationship (e.g., linear) produced a significant association between the Linda and Tom problems (r (118) = .28, p < .002), and a marginally non-significant correlation between the Jack and Linda problems (r (118) =

.18, $p < .052$).

Order Effects. Trend analyses were conducted to examine whether position of a problem within the set of five problems (e.g., whether a problem was the first or the last problem attempted) had an effect on responses. For each problem, linear, quadratic, cubic and quartic contrasts were calculated (Keppel, 1989) on absolute deviation scores. A significant linear trend was found for the Linda problem (F (1,113) = 9.01, MS_e = 2.20, $p < .01$). Examination of Table 7 reveals that subjects' absolute deviation scores were greater, not less, the further the Linda problem was presented within the set of five problems. There was also a significant cubic trend for the Tom problem (F (1,114) = 4.89, MS_e = 1.40, $p < .03$), and a significant quartic trend for the Jack problem (F (1,114) = 4.85, MS_e = 702.23, $p < .03$). The latter is attributable to smaller absolute deviation scores when the problem was the third or fifth problem presented.

Individual Difference Measures. Deviations from the normative response were associated with greater attributional complexity for the Tom problem (r (59) =.21, $p < .025$), but lesser attributional complexity for the Bret problem (r (58) = -.26, $p < .01$) which replicates Devine (1989). High self-monitoring was marginally, though non-significantly, associated with deviations from the normative response for the Bret problem.

Table 7

Mean Absolute Deviation Scores by Problem Position, Experiment 1.

	Position				
Problem	P1	P2	P3	P4	P5
Cab	31.3	33.4	31.7	32.6	34.3
Jack	42.2	41.3	26.4	42.2	32.2
Linda	3.9	4.0	4.6	4.7	4.9
Tom	4.7	5.3	4.7	4.5	4.8
Bret	4.0	4.2	4.6	4.2	4.3

Note. P1 is placement of the problem in position one, P2 is placement of the problem in position two, and so on.

Among the individual difference measures, Attributional Complexity and Need for Cognition scores were positively correlated (r (110) = .40, p < .0001). Somewhat unexpectedly, scores on the Psychological Research Survey were correlated with scores on both the Attributional Complexity (r (110) = .47, p <.0001) and Need for Cognition scales (r (110) = .35, p < .0002). On the other hand, scores on the Self-Monitoring scale were not correlated with scores on either the Psychological Research Survey (r (111) = -.05), the Attributional Complexity scale (r = (111) = -.08), or the Need for Cognition scale (r (112) = .06).

An additional individual difference variable considered was whether subjects who had completed at least one course in mathematics or logic (n = 77) would show smaller deviations from the normative response. For the absolute deviation scores, no significant differences were found between sophisticated and naive subjects for the Cab (F (1,117) = 0.05), Jack (F (1,116) = 0.00), Linda (F (1,114) = 1.25), Tom (F (1,116) = 0.42), or Bret problems (F (1,116) = 1.23).

Confidence Ratings. As a first step in analysis of subjects' confidence in their responses, correlations between absolute deviation scores and confidence ratings were calculated. These correlations were significant (p < .0001) for the Linda (r (116) = .36), Tom (r (117) = .51), and Bret problems (r (117) = .40), and significant at the p

< .01 level for the Jack problem (r (117) = .24). The correlation between the absolute deviation scores and confidence ratings was not significant for the Cab problem (r (118) = -.01). For all but the Cab problem, the more subjects deviated from the normative response, the greater was their confidence in the correctness of the response they had made.

To examine the factors affecting subjects' confidence in their judgments, confidence ratings served as the dependent variable in a 5 (problem) x 2 (gender) x 2 (normative sophistication) mixed ANOVA. There was a significant main effect for the within-subject factor of problem (F (4,444) = 13.93, MS_e = 1.46, p < .0001). Tukey HSD post-tests (p < .05) revealed subjects were significantly less confident in their response to the Cab problem (M = 4.50 on the seven-point scale) than in their responses to the Jack (M = 5.36), Linda (M = 5.19), Tom (M = 5.38) and Bret problems (M = 5.55). There were also significant main effects for gender (F (1,111) = 11.31, MS_e = 4.29, p < .001) and for normative sophistication (F (1,111) = 6.01, MS_e = 4.29, p < .02), with sophisticated and male subjects, respectively, being more confident in their responses.

There were no significant interactions between problem, sophistication, and gender. Males or sophisticated subjects may have appeared more confident because there were more

subjects who were both male and sophisticated. However, differences in the frequency of male/naive (n = 18), female/naive (n = 22), male/sophisticated (n = 41), and female/sophisticated subjects (n = 36) were not significant (X^2 (1,n=117) = 0.72).

Discussion

A normative interpretation of a base-rate problem requires a single solution be formally calculated, and it was hypothesized that interpreting a problem in such a manner would result in smaller deviations from the normative solution. Interpreting a problem as requiring a formal analysis resulted in smaller deviations from the normative response for the Jack, Linda, and Tom problems. Although none of the problems were seen to permit more than one correct answer, only the Cab and Jack problems were perceived to demand a formal analysis. Indeed, only subjects perceptions of the Cab problem coincided with the normative requirement of a formal analysis and a single correct solution.

A normative interpretation also regards base-rate information to be more relevant than likelihood information, which was hypothesized to be associated with smaller deviations from the normative response. Only for the Jack and Tom problems were subjects' endorsement of the base-rate information as relevant associated with smaller deviations from the normative response. At best, base-rate information

was seen by subjects to be no more relevant to solving the problems than likelihood information. Subjects apparently do not give baserates the emphasis either normative formulations (e.g., Bayes' theorem) or researchers dictate they should receive. Rather, it was the perceived relevance of the likelihood information that was positively associated with deviations from the normative solution.

The hypothesis that deviations from the normative solutions would be smaller the later a problem was presented in the set of five problems was not supported. Rather than this simple linear relationship between problem placement and deviation scores, trend analysis detected complex and non-linear relationships. It may be that placement of one problem relative to the position of some other problem is of consequence, rather than whether a problem is presented at the beginning or end of the set of problems.

The absence of significant relationships between the five problems was unexpected. On one hand, the problems were selected to represent a range of probability (e.g., conjunction fallacy) and social decision tasks (e.g., dispositional bias). However, all of the problems share the demand that case information be ignored, and only base-rate information be used. If deviations from normative solutions are the product of representative thinking, the absence of relationships between the problems suggest heuristics do not reflect a task-independent, problem solving style, but

rather are invoked by specific attributes of each problem. These attributes may be quite subtle, in so far as problems similar in structure and appearance (e.g., the Tom and the Jack problems) were unrelated.

Of the three individual difference variables considered in Experiment 1, only attributional complexity correlated with deviations from the normative solutions. A tendency toward making complex attributions was associated with deviations from the normative solution for the Tom problem. However, less complex or more simple attributions were associated with deviations from the normative solution for the Bret problem. The latter is consistent with Devine (1989), who surmised that attributionally complex individuals would recognize the absence of choice for Bret. Attributionally complex individuals might also be more strongly disposed to attempt to discern Tom's area of graduate study from the personal description provided. As an aside, significant correlations between the Psychological Research Survey and the two measures of cognitive complexity were unexpected. One substantive explanation for this relationship is that individuals who desire sophisticated causes for behavior and who possess a high level of engagement in cognitive activity might regard psychological research to be interesting and important.

With the exception of the Cab problem, the greater the deviations from the normative solutions, the more confident

subjects were in their responses. Subjects were significantly less confident in their responses to the Cab problem. Males and statistically sophisticated subjects, respectively, also exhibited greater confidence in their judgments. Because subjects' confidence ratings generally fell on the "Very Sure" side of the seven-point scale, differences in confidence ratings between problems, and between samples, reflected shifts from moderate and extreme confidence. Thus, it was the case that subjects who were "incorrect" in their responses to the problems were overconfident, rather than subjects who were "correct" in their responses were underconfident.

EXPERIMENT 2

Experiment 1 revealed base-rate problems to be a heterogeneous set of very difficult problems that were often interpreted differently than intended by researchers who subscribe to the normative model. However, individual difference measures were not found to be reliable sources for these alternative interpretations and the resulting deviations from the normative responses.

One plausible explanation for subjects' responses to these problems is that they may not correctly identify the appropriate base-rate and likelihood information. Correct identification would facilitate responding in a manner more consistent with the normative model. Although early theorists concluded error arose from overutilization of baserates (e.g., Peterson & Beach, 1967), according to Bar-Hillel (1980), "the genuineness, the robustness, and the generality of the base-rate fallacy" or the tendency to ignore base-rate information, is now an "established fact" (p. 215). However, one puzzling aspect of the base-rate fallacy is that outside the laboratory, baserates are often more influential than likelihood information. Nelson, et al., (1990), for example, have argued that stereotypes are a form of "everyday baserates" that are "very stable and difficult to eradicate" (p. 664). Their explanation for this contradiction between laboratory and real-world baserates is that the latter are strongly conditioned in

everyday encounters, while the former are conveyed by written material presented in a single experimental session.

It follows that subjects attempting laboratory problems may fail to incorporate baserates into their judgments because of their lack of salience. If the appropriate base-rate was identified and highlighted for subjects, for example by questions that highlighted the relevant information, such information might be incorporated into judgments, with the result being smaller deviations from the normative response.

In Experiment 2, one-third of the subjects were provided with a multiple-choice format question that cued identification of the appropriate value for the base-rate in each of the five problems. Another one-third of the subjects answered the standard version of each problem without a multiple-choice question preceding their judgment. It was predicted that subjects provided with the base-rate cue would deviate significantly less from the normative solution.

If the base-rate cue and standard version groups differ, it might be argued this difference arose not from making the base-rate more salient, but rather from an experimental demand to use the base-rate information. Thus, the remaining one-third of the subjects in Experiment 2 received <u>both</u> a base-rate and likelihood multiple-choice question prior to attempting each problem.

The saliency of base-rate information may also vary across base-rate problems. For the Cab and the Jack problems, baserates are explicitly stated (e.g., percentage of cabs in the city of one of two colours; number of lawyers and engineers in the population). For the Tom problem, the base-rate must be constructed by the subject (e.g., the number of education and library science graduate students in the subject's university). For the Linda and Bret problems, recognition of the appropriate base-rate requires intermediate judgments. In the case of Linda, subjects must recognize that the population of bank tellers who are feminists must be less than the population of bank tellers. In the case of Bret, subjects must recognize that Bret was assigned to write the counter-attitudinal essay, and that most students believe the level of support for AIDS research should be at least maintained or even increased. Because it was expected that there would be substantial differences between problems, all five problems were administered again in Experiment 2. Specifically, it was predicted that there would be a greater number of errors in base-rate identification for the Bret, Linda, and Tom problems, than for the Cab and Jack problems.

Method

Subjects. One-hundred and fourteen introductory psychology students from the subject pool at the University of Manitoba participated in Experiment 2. Subjects received

credit for their research participation.

Materials and procedure. The procedure was comparable to that of Experiment 1, with the addition of base-rate and likelihood cues in the form of multiple-choice questions. One-third of the subjects were not provided with any cues. One-third were provided with only base-rate multiple choice questions. The remaining third were provided with both base-rate and likelihood multiple choice questions. Subjects gave their response to each of the five problems presented in random order, indicated their confidence in their judgments, and provided relevance ratings for statements derived from the five problems.

Results

Condition Analysis. To examine differences between subjects provided with standard, base-rate cues only, and both base-rate and likelihood cues, a series of ANOVAs were conducted on absolute deviation scores and confidence ratings. All F tests were calculated with 2 and 111 degrees of freedom. Although an examination of the cell means reveals absolute deviation scores were smaller in the base-rate cue condition for all five problems, these effects were not significant for the Cab (F = .82), Jack (F = 1.10), Linda (F = 1.82), Tom (F = .35) and Bret (F = .81) problems (Table 8).

Accuracy. As can be seen in Table 9, subjects varied across problems in their accuracy for correctly identifying

Table 8

Mean Absolute Deviation Scores for Responses to the Five Judgment Problems by Condition, Experiment 2

Condition	Cab	Jack	Linda	Tom	Bret
Standard	30.92	28.89	4.58	4.74	4.58
Base-rate	29.89	26.84	3.95	4.47	4.18
Base-rate and Likelihood	33.24	36.05	4.16	4.52	4.47

Table 9

Multiple Choice Alternatives and Frequencies of Subjects Endorsing Alternatives, Experiment 2

Cab

The proportion of Green to Blue cabs in the city is approximately
a. 1:1 (2)
b. 3:1 (5)
c.* 6:1 (56)
d. 9:1 (13)

If the witness in the problem was asked ten times to identify cabs as either Blue or Green, the witness should be correct approximately _____ of the time.
a. 2/5 (0)
b. 3/5 (8)
c.* 4/5 (27)
d. all (0)

Jack

There are about _____ lawyers as engineers in the sample described above.
a. half as many (8)
b. the same number of (2)
c.* twice as many (60)
d. three times as many (6)

Jack is most likely to spend a Sunday afternoon
a. reading the Financial Post (1)
b. playing football (1)
c.* renovating his den (36)
d. driving in the country (0)

Linda

The proportion of female bank tellers at your bank you estimate are feminists is
a.* one-quarter (42)
b. one-half (17)
c. three-quarters (14)
d. all (2)

In college, Linda was most likely to subscribe to the magazine _____.
a. People (26)
b. Ladies Home Journal (7)
c.* Ms. (3)
d. Glamour (2)

Tom

For every 100 graduate students in education at this university, there are roughly _____ graduate students in library science.
a.* less than 25 (50)
b. about 50 (14)
c. the same number of (10)
d. more than 100 (2)

Tom is most likely to spend an evening _____.
a. at a professional hockey game (1)
b. volunteering at a crisis line (0)
c.* organizing a computer bulletin board (16)
d.* watching television (21)

Bret
Most students in your introductory psychology class probably believe that funding for research into AIDS should be _____.
a. decreased (5)
b. maintained at current levels (10)
c. increased somewhat (36)
d.* increased substantially (25)

One of the points Bret makes in his essay is that _____.
a. AIDS has not acquired media attention (1)
b. AIDS is a common disorder (2)
c. little energy has been invested into AIDS research (3)
d.* more money should be directed toward other diseases (32)

Note. The first question for each problem is the base-rate question (n = 38), the second question is the likelihood question (n = 76). Acceptable answer(s) are indicated by an asterisk. One subject failed to answer the likelihood question for the Cab problem, and one subject failed to answer the base-rate question for the Linda problem.

the appropriate base-rate and/or likelihood values. The proportion of subjects who made an error in identification of the appropriate base-rate was greatest for the Bret (67%), Linda (44%), and Tom problems (34%), and was least for the Cab (26%), and Jack problems (21%). There were no significant correlations between errors in base-rate identification and deviation scores.

Discussion

Prompting subjects to attend to base-rate values did not lead to significantly smaller deviations from the normative solutions to each problem. Thus, the base-rate fallacy cannot be attributed to a failure to identify the appropriate base-rate. Subjects made the greatest number of errors in identifying the appropriate base-rate information for problems where the base-rate information was not explicitly provided. However, these errors were not associated with the magnitude of deviations from the normative solutions. The latter result is consistent with the finding from the first experiment that baserates are seen by subjects to have little relevance to problem solving, and thus little impact upon problem solutions.

Consistent with the second hypothesis, the greatest number of errors in base-rate identification were for problems where the base-rate was not explicitly provided, specifically the Bret, Linda and Tom problems. In contrast, the Jack problem explicitly provides subjects with the

appropriate base-rate information, and across Experiments 1 and 2, the Jack problem resulted in the largest proportion of subjects providing the normative response. However, the Cab problem also explicitly provided subjects with the appropriate base-rate information, and only about one subject in a hundred provided the normative response to the Cab problem. Thus, base-rate explicitness is not a sufficient explanation for deviations from normative responses for these problems.

EXPERIMENT 3

The objective of the first two experiments was, in part, to ascertain how subjects' interpretations of the judgmental task affected their performance on base-rate judgment problems. Somewhat unanticipated findings from these experiments were the lack of correlations among the five problems, and the significant differences between problems on a number of dimensions. Because subjects' general perceptions of the experimental task were not predictive of their responses, an appropriate research strategy was to separately examine each problem in detail for its phenomenological correlates.

This detailed examination was made more meaningful by systematically varying each problem in a manner that affected the magnitude of deviations from the normative solution, and the phenomenological correlates of each version of each problem were then contrasted. Deviations from the normative solutions for base-rate problems have been consistently reduced by directly modifying either the base-rate or the likelihood statements. In the former case, Tversky and Kahneman (1982a) altered the Cab problem to make the base-rate information appear causal. In the latter case, Ginosar and Trope (1980) altered likelihood information in the Jack and Tom problems by rewriting the personality descriptions to include stereotypic elements from both professions or graduate majors. Along the same

lines, Margolis (1987) reported fewer conjunction errors for a version of the Linda problem that appeared to make the decision task less ambiguous, and Fein, et al., (1990) provided subjects with an ulterior motive for the behavior of an essay writer in an attitude attribution paradigm.

Although these manipulations have previously been shown to reduce deviations from normative solutions, why they are successful is not well understood. One set of explanations is that baserates are made causal (e.g., Tversky & Kahneman, 1982a), more relevant (e.g., Bar-Hillel, 1982), or more salient (e.g., Fein, et al., 1990). Alternatively, it may be the relevance of the likelihood information that is affected. A third possibility is that these modifications alter subjects' interpretation of what is demanded by the experimental task. Subjects may come to see the task in a manner that leads them to respond with the normative solution.

In order to permit a detailed examination of the phenomenological correlates of base-rate problems, and to contrast interpretations of standard and alternative versions of these problems, several new techniques were employed in addition to the PEQ measures from the first experiment. Subjects were asked to offer reasons for making ratings of high and low values for two of the five problems (Wright, Luus & Christie, 1990), and to endorse one of two interpretations of two other problems (Tversky & Kahneman,

1983).[8] The former required subjects to explain their response to a particular problem, and to consider why another subject might have responded differently. The latter provides for a direct test of whether subjects' endorse a normative or a non-normative interpretation of the solution to each problem.

Based on previous research, it was hypothesized that most subjects would report non-normative interpretations of the problems, and respond with substantial deviations from normative solutions. However, it was hypothesized that for alternative versions of the problems, deviations from the normative solution would be smaller, and that subjects would endorse a normative interpretation of the task.

It was also hypothesized that deviations from the normative solution would be smaller for subjects who offered an interpretation of the problem consistent with the normative interpretation. In other words, such subjects were expected to give a single correct solution obtained through formal analysis, with the base-rate seen as more relevant than the likelihood information.

Method

Subjects. One-hundred and twenty first-year introductory psychology students from the University of Manitoba subject pool participated in Experiment 3.

---

[8]Subjects were also asked to rewrite one problem in their own words (Scholz, 1987). Like Scholz, however, this procedure did not produce interpretable results.

Subjects received credit toward a research participation requirement.

Materials and Procedure. Subjects were assembled in groups of about thirty to complete the experimental booklet. Within each booklet were the same five judgment problems from Experiment 1. Half of the booklets contained the standard version of each problem that was presented to subjects in Experiment 1, and half contained the alternative version (Appendix C) previously found or predicted to reduce deviations from the normative response. Two of the manipulations directly altered statements leading to the base-rate information (the Cab and Bret problems), and two altered the likelihood information (the Jack and Tom problems).

The alternative version of the Cab problem was the 'causal' version (Tversky & Kahneman, 1982a) that implied the base-rate for accidents is a product of the different safety practices of the two cab companies; "Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs" (p. 157).

The alternative versions of the Jack and Tom problems were created by Ginosar and Trope (1980). The following is the alternative version of the Jack problem:

> [Jack] is a 45-year-old man. He is married and has four children. He is generally conservative,

careful, ambitious, competitive, and argumentative. He is interested in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles. (p. 233)

This version combines elements of both a stereotypic engineer (e.g., mathematical puzzles) and lawyer (e.g., interested in politics). [9]

The alternative version of the Tom problem included elements of a stereotypic education (e.g., enjoys interacting with others) and library science graduate student (e.g., need for order).

[Tom] has high intelligence and a strong drive for competence. He has a need for order and clarity and for neat and tidy systems in which every detail finds its appropriate place. His writing is rich and he expresses himself orally very well. He has much feeling and sympathy for other people and enjoys interacting with others. He has a deep moral sense and likes to give of himself to others. (Ginosar & Trope, 1980, p. 234)

The alternative version of the Bret problem suggested an external motive for the behavior of the author of the

---

[9]The presentation of the Jack problem in Experiment 1 did not include the information that Jack had been randomly selected from a sample of 30 engineers and 70 lawyers. Both the standard and alternative versions of the Jack problem in Experiment 2 included this information.

essay. Fein, et al., (1990) told subjects that the professor of the student author of an essay strongly supported the minority opinion. The alternative version of the Bret problem, specially constructed for this study, read "Bret was assigned to write a brief essay arguing for less funding for research into AIDS, an argument his professor had endorsed in class."

Finally, the alternative version of the Linda problem (Margolis, 1987) was created by Margolis to combat scenario effects.

> Linda is 31 year old, bright and outspoken. As an undergraduate she majored in philosophy and was active in the environmental and civil rights movements. A personnel survey showed that of clerical workers in banks (including tellers) fewer than 1% have personality profiles that sound similar to Linda's.
>
> If you stood to win $10 if the statement you choose turns out to be true (whether or not the other statement is also true), which choice is more likely to win you the $10.... (p. 166)

In addition to the PEQ measures from Experiment 1, one of two other techniques were used to investigate subjects' interpretations of each judgment problem. The choice of the particular technique for each problem (e.g., the second and fourth problem encountered to receive the "relevant reasons"

technique) was arbitrary but held constant across all subjects. Because problem order was randomized, each problem was followed by each of the techniques approximately the same number of times. Both techniques were not required for all five problems to lessen fatigue, to avoid a response set, and to reduce experimental demands. For the second and fourth problems, subjects were requested to write "relevant reasons, considerations, arguments, etc. that would justify recording a high number on the scale they had just completed" (Wright, et al., 1990, p. 266). The same task was then completed for a low number response.

For the first and third problems, subjects were asked to endorse either an interpretation consistent with the normative solution, or an interpretation that emphasized individuating information. Subjects responded on a seven-point scale, with the anchors 'Strongly Endorse Argument 1', the first interpretation presented, and 'Strongly Endorse Argument 2', the second interpretation presented. The two interpretations of the Cab problem were taken from Lyon and Slovic (1976), and of the Linda problem from Tversky and Kahneman (1983). The two interpretations of the Jack, Tom W., and Bret problems were constructed by the author for this study. Interpretations presented to subjects for each of the five problems are reproduced in Appendix C.

Confidence ratings were obtained for each problem, and order of presentation of problems was again examined. The

experimental booklet concluded with the PEQ from the first experiment.

Results

Responses to the Judgment Problems. Prior to testing the hypotheses of Experiment 3, the effects of alternative versions of the problems on deviations from the normative solutions for each of the five problems were examined. Summary statistics for responses to standard and alternative versions of the problems appear in Table 10. Responses to the standard version of each problem were comparable to those from the first experiment.

To test whether subjects' responses were closer to the normative solution for the alternative versions of the five problems, version of the problem was a between-subject factor in a series of ANOVAs. Absolute deviation scores were the dependent variables.[10] Subjects provided with alternative versions of the Jack (F (1,117) = 24.20, MSe = 523.34, $p < .0001$), Tom (F (1,117) = 90.65, MSe = 2.29, $p < .0001$) and Linda problems (F (1,117) = 4.03, MSe = 2.23, $p < .05$) produced significantly smaller absolute deviation scores. Smaller absolute deviation scores were not found for the Cab (F (1,117) = 0.22, MS_e = 97.54, $p < .65$), and Bret problems (F (1,117) = 1.42, MS_e = 1.91, $p < .23$).

---

[10] Gender and normative sophistication were also included as between-subject factors in this analysis. However, there were no significant main effects for gender or normative sophistication, nor any significant interactions.

Table 10

Summary Statistics for the Standard and Alternative Versions of the Five Judgment Problems, Experiment 3

	Problems				
Version	Cab	Jack	Linda	Tom	Bret
Standard					
M	54.78	69.25	5.50	5.85	2.58
S	28.52	25.76	1.20	1.18	1.40
Median	69	77.5	6	6	2.5
Mode	80	30	6	6	2
M_d	30.25	40.15	4.50	4.85	4.42
S_d	10.24	24.30	1.20	1.18	1.41
Alternative					
M	58.93	42.48	4.96	3.20	2.30
S	26.57	26.31	1.74	1.78	1.36
Median	70	30	6	3	2
Mode	80	30	6	2	2
M_d	31.06	19.78	3.97	2.20	4.70
S_d	9.42	21.28	1.74	1.78	1.36

Note. *n* = 60 for standard and alternative versions of all problems.

To test whether the absence of significant correlations between problems found in Experiment 1 was replicated in Experiment 3, correlations between the absolute deviation scores were calculated. For the standard versions of the problem, there were no significant correlations between the absolute deviation scores. For the alternative versions of the problems, there was a small, significant negative correlation between the Tom and Jack problems (r (59) = -.28, $p < .05$), and a small, significant positive correlation between the Tom and Bret problems (r (58) = .26, $p < .05$). Thus, subjects who deviated from the normative solution for the alternative version of the Tom problem also tended to deviate more from the normative solution for the Bret problem, but less from the normative solution for the Jack problem.

Solutions Demanded. It was hypothesized that alternative versions of the problems would lead to interpretations of the problems more consistent with the normative interpretation, and that this would, in turn, lead to smaller deviations from the correct solution. In order to determine whether this hypothesis was supported, the PEQ measures from Experiment 1 were examined. One set of PEQ measures asked subjects to ascertain the number of correct responses and the form of analysis the problems demanded. The first stage in the analysis was to determine whether alternative versions of the problems affected subjects'

responses on the number of correct solution and form of analysis items. The two versions of each problem (standard or alternative) was a between-subject factor, with problem as a within-subject factor.

For the number of correct responses permitted, neither version of the problem (F (1,115) = 0.08, MS_e = 9.17, p <.77) nor the version by problem interaction (F (4,460) = 1.79, MS_e = 2.06,p <.13) were significant. Thus, subjects provided with standard or alternative versions of the problems did not differ on their perceptions of the number of correct solutions demanded. Problem was a significant within-subject factor (F (4,460) = 2.91, MS_e =2.06, p <.02). The Cab problem (M = 3.18) was seen as more demanding of a single solution compared to the Linda problem (M = 3.70). The Jack (M = 3.43), Tom (M = 3.68), and Bret problems (M = 3.64) did not differ.

For the form of analysis required, there was a significant main effect for version (F (1,116) = 4.33, MS_e = 3.58, p <.04). For the Jack (F (1,117) = 2.83, MS_e = 3.46, p <.10), Linda (F (1,117) = 4.33, MS_e = 2.39, p <.04), and Bret problems (F (1,117) = 3.22, MS_e = 2.52, p <.075), alternative versions of the problems were associated with a more formal analysis. The interaction of version and problem was not significant (F (4,464) = 1.73, MS_e = 2.32, p <.14). Again, there was a significant within-subject effect for problem (F (4,464) = 67.91, MS_e = 2.32, p <.0001).

Subjects perceived the Cab problem to be more demanding of a formal analysis (M = 5.49) than the Jack problem (M = 3.85), and both to be significantly more demanding of a formal analysis than the Linda (M = 2.87), Tom (M = 2.85) and Bret problems (M = 2.91).

In order to test the possibility that these perceptions affected absolute deviation scores, the median split procedure described in Experiment 1 was used to convert responses to these two PEQ items into between-subject factors for a series of ANOVAs. This analysis was done separately for standard and for alternative versions of the problems. No significant main effects for number of correct responses or form of analysis for either standard or alternative versions of the problems were found.

Relevance of Problem Statements. A robust finding from Experiment 1 was that perceived relevance of base-rate and likelihood statements strongly predicted deviation scores. In Experiment 3, a series of mixed-model ANOVAs with problem version (standard or alternative) as a between-subject factor, and the problem statements as a within-subject factor, served to evaluate the effect of problem version on subjects' perceptions of the relevance of base-rate and likelihood problem statements. Because only base-rate and likelihood statements were considered in this analysis, the Linda problem was excluded.

A summary of the results of the ANOVAs are presented in

Table 11. Little support was found for the proposition that alternative versions of the problem lead subjects to view the base-rate statements as more relevant. Instead, it was the relevance ratings of the likelihood statements, if at all, that were affected by manipulation of elements for several of the problems. For the alternative version of the Cab problem (Table 12), the likelihood statement was judged significantly more relevant than the base-rate statement, even though the latter was modified to be causal. The base-rate and likelihood statements were seen as equally relevant for the standard version of the problem, although the likelihood statement for the standard version was judged to be significantly more relevant than the likelihood statement for the alternative version (F (1,117) = 7.79, MS_e = 1.83, p < .006). The base-rate statements for the two versions did not differ in their relevance ratings, and only the likelihood statement of the standard version correlated with deviation scores.

For the Jack problem (Table 13), the base-rate statement was judged to be significantly less relevant than two of the likelihood statements for the standard version, and one of the likelihood statements for the alternative version. The addition of the competitive and argumentative elements to the first likelihood statement of the alternative version resulted in it being judged significantly more relevant (F (1,115) = 21.87, MSe = 2.76,

Table 11

ANOVA Results for Relevance Ratings of Statements from Standard and Alternative Versions of the Five Problems, Experiment 3

Problem	df	MS_e	F	p
Cab				
Version	1,117	2.44	0.51	.48
Statements	1,117	2.92	15.35	.0002
VxS	1,117	2.92	6.08	.015
Jack				
Version	1,115	6.29	0.19	.66
Statements	3,345	2.79	10.81	.0001
VxS	3,345	2.79	9.37	.0001
Tom				
Version	1,116	3.19	11.39	.001
Statements	3,348	2.33	43.36	.0001
VxS	3,348	2.33	6.82	.0002
Bret				
Version	1,117	4.72	0.21	.65
Statements	3,351	2.83	3.35	.02
VxS	3,351	2.83	0.24	.87

Table 12

Relevance Ratings of Statements in the Cab Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 3

Statement	M	r
Standard		
85% of the cabs in the city are Green and 15% are Blue. [BR]	5.17[a]	-.25
The witness made correct identifications in 80% of the cases and erred in 20% of the cases. [LH]	5.49[a]	.37**
Alternative		
Although the two companies are roughly equal in size, 85% of the accidents in the city involve Green cabs and 15% involve Blue cabs. [BR]	4.77[b]	-.22
The witness made correct identifications in 80% of the cases and erred in 20% of the cases. [LH]	6.18[a]	.19

Note. [BR] = Base-rate element, [LH] = Likelihood element. For each version, mean ratings with different superscripts significantly different according to Tukey HSD post-hoc tests ($p < .05$).

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

Table 13

Relevance Ratings of Statements in the Jack Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 3

Statement	M	r
Standard		
30 engineers and 70 lawyers, all successful. [BR]	3.83[a]	-.26*
He is generally conservative, careful and ambitious. [LH1]	3.55[a]	.36**
He shows no interest in political and social issues. [LH2]	5.22[b]	.50****
He spends most of his free time on his many hobbies: home carpentry, sailing, and mathematical puzzles. [LH3]	5.15[b]	.42***
Alternative		
30 engineers and 70 lawyers, all successful. [BR]	3.96[a]	-.16
He is generally conservative, careful and ambitious, competitive and argumentative. [LH1]	5.00[b]	.20
He is interested in political and social issues. [LH2]	4.70[ab]	.11
He spends most of his free time on his many hobbies: home carpentry, sailing, and mathematical puzzles. [LH3]	4.60[ab]	.39**

Note. [BR] = Base-rate element, [LH1] = Likelihood element

1, and so on. Mean ratings with different superscripts significantly different for each version of the problem according to Tukey HSD post-hoc tests ($p < .05$).

* $p < .05$. ** $p < .01$. *** $p < .001$ **** $p < .0001$

$p < .0001$), but did not affect the relevance of the base-rate statement. Endorsing the likelihood statements as relevant also correlated positively with deviations scores for the standard version of the problem.

Examination of the results for the Tom problem (Table 14), revealed that contrary to a normative interpretation of the problem, the base-rate statement was judged to be significantly less relevant than the likelihood statements for both standard and alternative versions. Reference to Tom's expressiveness (F (1,116) = 6.32, MS_e = 1.98, $p < .01$), and giving of himself to others (F (1,116) = 26.38, MS_e = 2.66, $p < .0001$, in the alternative version, resulted in those statements being judged as significantly more relevant. However, relevance ratings did not correlate with deviations from the normative solution.

Evaluating standard and alternative versions of the Bret problem separately, no statement was judged to be more nor less relevant than any other statement (Table 15). Relevance ratings correlated positively with deviations from the normative response, but for the alternative rather than for the standard version of the problem.

Interpretation of the Problem. For the first and third problem presented in Experiment 3, subjects were asked to examine a normative or base-rate interpretation and a non-normative or likelihood interpretation of solutions for each problem (Appendix C), and indicate on a seven-point scale

Table 14

Relevance Ratings of Statements in the Tom Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 3

Statement	M	r
Standard		
The following is a description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school. [BR]	3.64[a]	.01
His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and flashes of imagination. [LH1]	4.93[bd]	.05
He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. [LH2]	5.64[cd]	-.01
Self-centered, he nonetheless has a deep moral sense. [LH3]	4.21[ab]	-.01
Alternative Version		
The following is a description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school. [BR]	3.50[a]	.10
His writing is rich and he expresses himself orally very well. [LH1]	5.58[b]	.07

He has much feeling and sympathy for other people and enjoys interactions with others. [LH2]	5.80[c]	-.06
He has a deep moral sense and likes to give of himself to others. [LH3]	5.75[c]	-.09

Note. [BR] = Base-rate element, [LH1] = Likelihood element 1, and so on. Mean ratings with different superscripts significantly different for each version according to Tukey HSD post-hoc tests ($p < .05$).

* $p < .05$.

Table 15

Relevance Ratings of Statements in the Bret Problem and the Correlation between the Statements and Absolute Deviation Scores, Experiment 3

Statement	M	r
Standard		
Bret was assigned to write a brief essay arguing for less funding for research into AIDS. [BR]	5.19[a]	-.20
Bret wrote that AIDS is comparatively rare and did not warrant the coverage in the media it has received. [LH1]	4.68[a]	.27
Bret wrote that an enormous amount of energy and costly research is directed toward the investigation of AIDS at the expense of more important life-threatening diseases. [LH2]	5.08[a]	.20
Bret wrote that if the same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit. [LH3]	4.92[a]	.19
Alternative Version		
Bret was assigned to write a brief essay arguing for less funding for research into AIDS, an argument that his professor had endorsed in class. [BR]	5.27[a]	-.19

Bret wrote that AIDS is comparatively rare and did not warrant the coverage in the media it has received. [LH1]	4.41[a]	.39**
Bret wrote that an enormous amount of energy and costly research is directed toward the investigation of AIDS at the expense of more important life-threatening diseases. [LH2]	4.91[a]	.19
Bret wrote that if the same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit. [LH3]	4.90[a]	.37**

Note. [BR] = Base-rate element, [LH1] = Likelihood element 1, and so on. Mean ratings with different superscripts significantly different for each version according to Tukey HSD post-hoc tests ($p < .05$).

** $p < .01$

the strength of their endorsement of one of the two interpretations. These endorsement ratings served as the dependent variable in a series of ANOVAs, with version of the problem (standard or alternative) as a between-subject factor. It was predicted that alternative versions of the problems would lead to significantly more normative interpretations. This was true for the Jack (F (1,46) = 4.18, MS_e = 2.87, p <.05), and Tom problems (F (1,46) = 6.08, MS_e = 3.29, p < .02), marginally nonsignificant for the Cab problem (F (1,45) = 3.66, MS_e = 3.63, p <.06), and non-significant for the Linda (F (1,45) = 1.08, MS_e = 2.02, p < .30) and Bret problems (F (1,47) = 0.16, MS_e = 2.98, p < .69).

Subjects provided with the alternative version of the Cab problem (M = 3.5 on the seven-point scale) leaned toward a likelihood interpretation of the solution, while subjects provided with the standard version of the Cab problem leaned toward a base-rate interpretation (M = 4.6). For the Jack problem, subjects who received the standard version (M = 2.5) more strongly endorsed the likelihood interpretation than did subjects who received the alternative version (M = 3.5). For the Tom problem, standard version subjects (M = 2.6) more strongly endorsed a likelihood interpretation, with alternative version subjects falling between the two interpretations (M = 3.9). In contrast, subjects endorsed the respective normative interpretations of the Linda (M =

3.1) and Bret problems ($\underline{M}$ = 5.1). Note that for the Linda problem, a low rating reflects the normative interpretation. The reverse is true for all other problems.

For the Cab, Jack, Linda and Tom problems, endorsing the non-normative interpretation was correlated with deviations from the correct response (Table 16). This effect was more pronounced for the alternative version of each problem.

Rationale for a High/Low Response. For the second and fourth problems presented, subjects were asked to give a rationale for a high and for a low response. The frequency of responses of various types are presented in Table 17. When a subject gave both a normative and non-normative rationale for the same problem, both categories were awarded a score. Few subjects produced base-rate rationales even when their response was close to the normative solution. The Cab problem produced the greatest number of base-rate responses. The numerous Other/No Responses for low numbers (high numbers for the Bret problem) reflected the difficulty subjects had in understanding how the problem could be responded to from a different perspective. The small number of base-rate responses precluded separate analysis for standard and alternative versions of the problems.

Confidence Ratings. In order to see if the effects for confidence ratings from Experiment 1 would replicate, a

Table 16

Correlations between Endorsement Ratings, Absolute Deviations from the Normative Judgment, and Confidence Ratings, Experiment 3.

	Problem				
Version	Cab	Jack	Linda	Tom	Bret
Standard					
d	-.41*	-.35	-.67***	-.36	.02
Conf.	-.11	-.16	-.40*	.06	-.21
Alternative					
d	-.65***	-.54**	-.82****	-.36	-.31
Conf.	-.02	-.22	-.06	.20	-.33

* $p < .05$. ** $p < .01$. *** $p < .001$. **** $p < .0001$

Table 17

Frequency of Reasons for High and Low Ratings of the Five Judgment Problems, Experiment 3.

	Cab	Jack	Linda	Tom	Bret
High Number					
Base-rate	10	4	5	0	10
Likelihood	35	31	39	40	14
Other/No Response	7	14	9	8	24
Low Number					
Base-rate	23	7	6	0	7
Likelihood	12	25	11	28	25
Other/No Response	19	18	32	20	17

Note. A low number represents a normative response for all problems except Bret.

mixed-model ANOVA examined effects of problem version, gender, and normative sophistication as between-subject factors, and problem as a within-subject factor on confidence ratings. These were variables predicted to affect confidence ratings in Experiment 1. The only significant between-subject effect was for problem version (F (1,110) = 4.49, MS_e = 189.15, p < .04). Subjects given the standard version of the Linda problem were somewhat more confident in their judgment (M = 5.25) compared to subjects given the alternative version (M = 4.80). There was also a significant within-subject effect (F (4,440) = 5.72, MS_e = 1.56, p <.0002). Subjects were significantly less confident in their responses to the Cab problem (M = 4.81) than in their responses to the Jack (M = 5.31), Tom (M = 5.32) and Bret (M = 5.53) problems, and less confident in their response to the Linda problem (M = 5.03) than in their responses to the Bret problem. No other main effects or interactions were significant.

Discussion

Alternative versions of the Jack, Linda, and Tom problems, but not the Bret and Cab problems, lead to responses significantly closer to the normative solutions. Although the alternative version of the Bret problem was patterned after a manipulation employed by Fein, Hilton and Miller (1990) in their attitude attribution study, the fact that the Bret problem was created specifically for this

study may explain why the manipulation was unsuccessful. In contrast, the alternative version of the Cab problem was identical to a version of the problem presented to subjects by Tversky and Kahneman (1982a) that was shown to reduce deviations from the correct solution in their study. This version of the Cab problem did not affect responses in this study, and pilot studies also found the alternative version of the Cab problem to be ineffective.

Standard or alternative version of a problem had little impact on subjects' perceptions of either the number of solutions or form of analysis. Once again, the Cab and Jack problems were seen to be more consistent with a formal analysis, and the Cab problem was strongly identified with a single correct response. Unlike the first experiment, form of analysis was not related to absolute deviation scores, and was related to confidence ratings for only standard versions of two of the five problems.

The perceived relevance of likelihood statements, but not base-rate statements, were affected by presentation of alternative versions of the problems. The addition of contradictory passages to the likelihood statements of the Tom and the Jack problems resulted in significantly smaller deviations from normative solutions. However, these modifications resulted in the likelihood information being judged more, not less relevant. There were no significant correlations between likelihood statements and absolute

deviation scores for alternative versions of the Tom and Jack problems. This outcome suggests the addition of passages inconsistent with the stereotypes of library science majors and engineers, respectively, reflects the reduction in relevance for the respective likelihood statements. Neither modification of the base-rate statement in the Cab problem to make it appear causal, nor the addition of a passage suggesting Bret was endorsing the position of his instructor, affected either the perceived relevance of the baserates or the magnitude of the deviation from the normative response.

One task required of subjects in Experiment 3 not required in the first experiment was to endorse one of two interpretations of each problem. Subjects presented with Standard versions of the Jack and Tom problems, and to a lesser extent the Cab problem, strongly endorsed non-normative explanations for their responses. In contrast, these same subjects endorsed normative interpretations for the Bret and the Linda problems. After being presented with the normative explanation for the Bret and Linda problem, some subjects spontaneously exhibited a sort of "I should have known it all along effect". This capitulation was not found for the Jack and Tom problems. For these problems, subjects were adamant that their interpretation was the correct one.

Comparing the two versions of each problem for

endorsement of normative and non-normative interpretations resulted in differences between the standard and alternative versions of the Jack and Tom problems, but not the Cab, Linda, and Bret problems. Subjects endorsed non-normative interpretations for the Jack and Tom problems, non-normative interpretations of the Linda and Bret problem, and were split between the two interpretations for the Cab problem. These findings parallel the results from the relevance judgments. More importantly, however, endorsement ratings were strongly associated with the magnitude of deviations from the normative solutions for both standard and alternative versions of the problems.

EXPERIMENT 4

The third experiment found some success in reducing deviations from the normative solution by altering likelihood information. This approach to remediating bias has limited application in the real-world because information that serves to disconfirm an established stereotype is not easily obtained and not actively sought by decision makers. A different approach to remediating bias follows from research on the 'frame' (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981) or problem solving set (Scholz, 1987) subjects operate under while trying to solve judgment problems. Researchers have attempted to assess and classify the sets within which subjects' approached the decision task.

On the basis of analysis of verbal protocols, Konold (1989) differentiated outcome-oriented and frequentist styles of solving judgment problems. Similarly, Scholz (1987) coded verbal protocols on an analytic/intuitive dimension, and Barnes (1985) identified applications of aleatory and epistemic logic to judgment problems. Outcome-oriented, intuitive and epistemic dimensions reflect a non-normative set that accepts likelihood information as relevant. In contrast, frequentist, analytic and aleatory dimensions reflect a normative set that regards only the

base-rate information as relevant.[11]

The idea of a general problem solving set has lead to attempts (e.g., Fischhoff, et al., 1979; Fischhoff & Bar-Hillel, 1984b) to manipulate subjects' frames or approaches to the experiment. These frames, in turn affect subjects' interpretation of the judgment problems. The implicit goal of this approach has been to seek a means to eliminate biased judgments without altering the content of the problem as was done in Experiment 3. For the most part, judgmental biases have been resistant to such efforts. For example, attempts to motivate subjects through monetary incentives and instructions to try harder have been unsuccessful (Fischhoff, 1982).

Fischhoff, et al., (1979) suggested base-rate information is ignored either because subjects fail to see its relevance to the problem or because subjects "lose [the base-rate] in the cognitive shuffle of combining various pieces of information" (p. 340). On the basis of the former interpretation, Fischhoff et al., hypothesized that increasing the salience of the base-rate would result in less error. Subjective Sensitivity Analysis (SSA) requires subjects to respond to the same problem a number of times, each time with a different base-rate specified. Although

---

[11]Bayes' theorem dictates that both base-rate and likelihood information be integrated. However, Gigerenzer and Murray (1987) point out that many base-rate problems demand the likelihood information be ignored, and only the base-rate information be considered.

Fischhoff et al., found subjects' responses to be sensitive to change in the base rates, the amount of adjustment was insufficient, and SSA training did not generalize to a second problem of the same general type.

A second series of experiments (Fischhoff & Bar-Hillel, 1984b) compared SSA to Isolation Analysis, in which subjects were asked to consider their response if only a selected piece of information was available, to Minimal Focusing, in which subjects were simply instructed to consider all items of information before responding, and to Balanced SSA, in which subjects were presented with various base-rate and likelihood values. For all four techniques, responses were somewhat closer to the normative response than what had been reported in previous studies. However, performance on subsequent base-rate problems suggested that these techniques had not "created additional understanding nor evoked latent understanding regarding the role of base rates" (p. 187).

It is somewhat surprising that the effects found for manipulating the experimental frame have not been more robust. Framing effects have been shown to be extremely powerful in other decision making contexts. For example, the original Kahneman and Tversky (1984) demonstration that manipulated whether the choice of one of two treatment programs was framed in terms of people living or people dying resulted in a reversal in the proportion of subjects

choosing one program over another. However, Fischhoff et al., (1979), and Fischhoff and Bar-Hillel (1984b) failed to include a comparison group that did not receive such instructions. Without such a comparison group, it is difficult to say that under those particular experimental circumstances, the manipulations resulted in smaller deviations from the normative solution than what might otherwise have been found. Most importantly, none of the previous studies systematically increased the explicit 'counterdemands' to use base-rate and ignore individuating information, nor examined the phenomenological correlates to these instructions. Without doing so, it would not be possible to arrive at any definitive answer as to why such manipulations are ineffective.

In Experiment 4, a variety of pre-experimental instructions were applied in an attempt to induce subjects to interpret the set of problems in a manner more consistent with the normative model. Phenomenological correlates to these instructions were assessed through the techniques employed in the preceding experiments. It was hypothesized that the stronger the counterdemands, the less responses to the five problems would deviate from the normative solution, and the more closely would responses on the phenomenological measures approximate the normative interpretation for the problems.

Method

Subjects. Two-hundred and forty first-year introductory psychology students from the University of Manitoba subject pool participated in Experiment 4. Subjects received credit toward a research participation requirement.

Materials and procedure. Subjects were assembled in groups of no more than thirty to complete an experimental booklet containing five judgment problems. Before attempting the problems, subjects received one of five forms of pre-experimental instructions:

(1) Standard Frame. The standard instructions provided to subjects in the first three experiments were employed in this frame.

(2) Aleatory Frame. For this group, instructions were patterned after a description of aleatory logic provided to judges of verbal protocols (Barnes, 1985).

> Aleatory logic requires thinking about objects or events with a focus on the objects or events as part of a larger set. Aleatory logic requires that unique characteristics of an object or event be ignored, and only the membership of the object or event in a set be considered in making judgments. Knowledge about individual elements is irrelevant.

(3) Demand-control Frame. These instructions were adapted from Wright and Wells (1988), and Wright et al., (1990).

These authors argued subjects avoid giving obvious answers to problems and believe that information provided by experimenters must be relevant to problem solutions. Subjects in this condition were told:

> The questions pertaining to some of the problems may not have obvious, straightforward answers. Some problems may not include enough information to answer some of the questions. Moreover, some problems may contain some information that is not relevant to the problem. You must decide on your own how relevant particular items of information are to the various questions asked.

(4) Modified Minimal Focusing Frame. The instructions given to subjects in this condition were patterned after Fischhoff and Bar-Hillel's (1984b) instructions to subjects attempting the Cab problem to consider information about both the witness and the number of cabs:

> For all of the problems you will encounter in this booklet, consider both the information about the characteristics of the specific individual or object and the information about the population or larger group that the individual or object was drawn from.

(5) Explicit Frame. These instructions simply told subjects to ignore likelihood information:

> In each of the following set of problems, you will

be provided with general information about some group of objects or people and specific information about one object or person. Ignore the specific information when you make judgments.

One of these framing instructions was presented to each subject before they attempted to solve the five judgment problems. In all other respects, this experiment replicated the procedure from Experiment 3 except that only the standard version of each problem was given.

Results

Responses to the Judgment Problems. Summary statistics for responses to the judgment problems, collapsed across framing instructions, are presented in Table 18. Because the only significant correlation between absolute deviation scores for the five problems was between the Bret and Tom problems (r (238) = .19, p < 005), each problem was examined individually to test each of the hypotheses.

To test the hypothesis that smaller deviations from the normative solutions would be found with non-standard problem frames, a series of contrasts were done between the Standard and the other four frame instructions. The contrast between the Standard and Demand frames was significant for the Jack problem (F (1,231) = 5.04, MS_e = 688.53, p <.03), and the contrast between Standard and Focusing frames was significant for the Tom problem (F (1,232) = 8.56, MS_e =

Table 18

Summary Statistics for the Five Judgment Problems, Experiment 4

	Cab	Jack	Linda	Tom	Bret
M	59.43	66.28	5.40	5.75	2.55
S	27.82	28.04	1.35	1.32	1.35
Median	75	75	6	6	2
Mode	80	30	6	2	2
M_d	32.01	37.52	4.4	4.75	4.44
S_d	11.14	26.33	1.35	1.32	1.35

Note. ns varied between 46 and 48 according to the number of missing responses.

1.70, $p < .004$), and for the Bret problem (F (1,232) = 7.45, MS_e = 1.78, $p < .007$). Cell means are presented in Table 19. In all three cases, the Standard frame was associated with significantly greater deviations from the normative solution. There were no significant main effects for gender or normative expertise and no significant interactions.

Confidence Ratings. Confidence ratings are an important measure of subjects' decision making process because a poor decision made with little confidence will have different consequences than the same poor decision made with publicly or privately expressed certainty (Dunning, Griffin, Milojkovic & Ross, 1990). Confidence ratings have also often been found to be poorly calibrated to accuracy (Lichtenstein, Fischhoff, & Philips, 1982), thus providing a unique measure of between-subject effects. Confidence ratings were also shown in Experiment 1 to be sensitive to individual difference variables such as gender and normative sophistication. For these reasons, a series of ANOVA's examined the effects of framing instructions and gender on confidence ratings (Table 20). Confidence ratings by problem, framing instructions, and gender are presented in Table 21. Tukey post-hoc tests ($p < .05$) revealed significantly greater confidence for the Modified frame compared to the Explicit frame for the Tom and Bret problems. Male subjects were significantly more confident than female subjects for all but the Linda problem.

Table 19

Mean Absolute Deviation Scores by Framing Instructions for Each Judgment Problem, Experiment 4.

	Problem				
Frame	Cab	Jack	Linda	Tom	Bret
Std	32.3	29.7	4.4	4.4	4.1
Alt	32.2	38.2	4.5	4.8	4.4
DC	33.4	41.7	4.2	4.8	4.5
MMF	30.6	39.0	4.6	5.1	4.9
Exp	31.6	38.9	4.3	4.6	4.3

Note. Std = Standard frame, Alt = Aleatory frame, DC = Demand frame, MMF = Modified frame, Exp = Explicit frame.

Table 20

Significant Test of Differences in Confidence Ratings by Framing Instructions and by Gender, Experiment 4.

Problem	df	MS_e	F	p
Cab				
Frame	4,225	2.44	0.38	.83
Gender	1,225		5.96	.02*
FxG	4,225		0.21	.93
Jack				
Frame	4,225	1.63	0.77	.55
Gender	1,225		13.12	.0004*
FxG	4,225		1.25	.29
Linda				
Frame	4,225	1.94	1.96	.10
Gender	1,225		1.12	.29
FxG	4,225		2.24	.06
Tom				
Frame	4,225	1.73	2.85	.02*
Gender	1,225		3.26	.07
FxG	4,225		2.12	.08
Bret				
Frame	4,225	2.08	2.75	.03*
Gender	1,228		9.89	.002*
FxG	4,228		2.18	.07

* $p < .05$

Table 21

Mean Confidence Ratings by Framing Instructions and by Gender, Experiment 4.

	Problem				
Frame	Cab	Jack	Linda	Tom	Bret
Standard					
Females	4.4	4.9	4.9	4.9	5.2
Males	4.8	5.8	4.8	5.5	5.6
Combined	4.7	5.4	4.8	5.3	5.4
Aleatory					
Females	4.7	4.9	5.4	5.3	4.7
Males	4.6	6.0	5.6	6.1	6.0
Combined	4.7	5.6	5.5	5.7	5.4
Demand					
Females	4.5	5.6	5.5	5.7	5.6
Males	5.0	5.7	5.0	5.1	5.7
Combined	4.8	5.6	5.1	5.3	5.6
Modified					
Females	4.4	5.5	4.9	5.8	5.9
Males	5.1	6.0	5.9	6.0	5.8
Combined	4.8	5.8	5.5	5.9[a]	5.8[a]
Explicit					
Females	4.2	5.2	4.6	4.9	4.4
Males	4.9	5.9	4.9	5.3	5.6
Combined	4.5	5.5	4.7	5.1[b]	5.0[b]

Note. Column means with different superscripts significantly different from each other ($p < .05$), Tukey post-hoc HSD test.

but the Linda problem.

Consistent with previous findings, there was a significant within-subject ANOVA for confidence ratings collapsed across framing instructions and gender (F (4,900) = 15.35, MS_e = 1.73, p <.0001). Tukey HSD tests revealed that subjects were less confident in their response to the Cab problem (M = 4.71) than in their response to the Jack (M = 5.60), Linda (M = 5.14), Tom (M = 5.45) and Bret (M = 5.46) problems. Subjects were also significantly less confident in their response to the Linda problem compared to the Jack, Tom and Bret problems.

Solutions Demanded. The effects of different framing instructions on the phenomenological measures were evaluated to test the hypothesis that these frames would affect subjects' interpretations of the problems, specifically whether more than one correct response for a problem was accepted, and whether an intuitive analysis to solve a problem was required. The number of correct responses permitted and the form of analysis (formal or intuitive) items were examined through a series of mixed-model ANOVAs. Framing instructions served as a between-subject factor, with problem as a within-subject factor. For number of correct responses, neither framing instructions (F (4,231) = 1.00, MS_e = 9.34, p <.41) nor the interaction of framing instructions by problem (F (16,924) = 1.41, MS_e = 2.73, p <.13) were significant. Problem was a significant within-

subject factor (F (4,924) = 13.90, MSe = 2.73, p <.0001). Consistent with the results from the earlier studies, the Cab problem (M = 3.44) was seen as more demanding of a single solution compared to the Linda (M = 4.44), Tom (M = 3.88) and Bret problems (M = 4.06), the Jack problem (M = 3.52) was seen as more demanding of a single correct response compared to the Linda and Bret problems, and the Tom problem was seen as more demanding of a single response compared to the Bret problem.

Examining subjects' perceptions of the form of analysis each problem demanded, there was no significant main effect for framing instructions (F (4,231) = 0.25, MS_e = 5.18, p <.91), nor for the interaction of framing instructions and problem (F (16,924) = 0.94, MS_e = 2.88, p <.52). There was a significant within-subject effect for problem (F (4,924) = 94.48, MS_e = 2.88, p <.0001). Again consistent with previous analyses, subjects perceived the Cab problem (M = 5.49) to be significantly more demanding of a formal analysis than either the Jack (M = 3.77), Linda (M = 2.74), Tom (M = 3.12) and Bret problems (M = 3.17), and the Jack problem to be significantly more demanding of a formal analysis than the Linda and Bret problems.

Relevance of Statements. In the same vein, a series of ANOVAs examined whether the different framing instructions affected subjects' ratings of whether base-rate or likelihood information was perceived to be relevant to solving the judgment problems. There was no significant

main effects for framing instructions for any of the base-rate or likelihood statements for any of the five problems.

Interpretation of the Problem. For the first and third problems presented, rather than for all five problems so as to reduce subject fatigue and experimental demands, subjects were asked to endorse one of two interpretations for each problem. These ratings served as a dependent variable in a series of ANOVAs, with framing instructions as the independent variable. Framing instructions were not significant for the Cab (F (4,89) = 0.21, MS_e = 3.51, p <.93), Jack (F (4,90) = 1.03, MS_e = 3.27, p < .40), Linda (F (4,90) = 1.26, MS_e = 2.69, p <.29), and Bret problems (F (4,89) = 1.04, MS_e = 3.19, p < .39). Framing instructions were significant for the Tom problem (F (4,91) = 2.87, MS_e = 2.33, p < .03). Subjects given modified instructions endorsed an interpretation consistent with the likelihood information (M = 1.45) more so than subjects given explicit instructions (M = 2.85).

Salience of the Framing Instructions. One possible explanation for the few significant effects for framing instructions on absolute deviation scores, confidence ratings, and phenomenological measures, is that subjects simply failed to attend to the instructions. To test this possibility, 208 of the 240 subjects in Experiment 3 were provided the five framing instructions at the conclusion of the experiment, and were simply asked to identify which of

the five instructions had been presented at the beginning of the experiment. A large number of Demand (n = 17) and Modified frame subjects (n = 15) failed to correctly identify the instructions they had read at the beginning of the session (Table 22). However, analyzing the results without these subjects did not alter the conclusions of the preceding analyses.

Discussion

Manipulation of pre-experimental instructions to subjects had little effect on their responses to the five judgment problems, their confidence ratings, or their interpretation of the problems. Subjects were somewhat less confident in their responses under non-standard frames, but these instructions had no effect on subjects' perceptions of the number of correct solutions, form of analysis, or relevance ratings. Confidence ratings tended to be a more sensitive measure than deviation scores because a manipulation that may lack the potency to affect judgments may still be capable of affecting the confidence of the decision maker in the judgment. Thus, Experiment 4 replicated previous studies that have found, at best, weak effects for instructions to remediate biased responding (e.g., Fischhoff, 1982).

The most parsimonious explanation for these failures is that subjects, even though told to pay close attention,

Table 22

Frequency of Responses Comparing Framing Instructions Assignment to Recognition Test Performance, Experiment 3

	Assigned				
Response	Std	Alt	DC	MMF	Exp
Std	33	0	0	0	0
Alt	2	38	4	1	3
DC	2	0	24	3	2
MMF	2	0	2	28	0
Exp	4	3	11	11	35
n	43	41	41	43	40
A	10	3	17	15	5
B	23.2	7.3	41.5	34.9	12.5

Note. Std = Standard frame, Alt = Aleatory frame, DC = Demand frame, MMF = Modified frame, Exp = Explicit frame. *n* = the number of subjects assigned to each condition and tested for recall; A is the number of subjects making an incorrect recognition response; B is the percentage of the number of subjects making an incorrect response (A - *n*).

simply failed to attend to the instructions. This explanation received support in Experiment 4 from the large number of subjects who failed the recognition test for instructions provided at the beginning of the experiment. One might conclude that this failure to attend to the instruction resulted from subjects being engaged in "mindless" information processing (Langer, Blank, & Chanowitz, 1978). A more charitable explanation for their poor performance on the recognition test was interference from the subsequent judgment task. A number of subjects post-experimentally reported having either forgotten or not understanding the instructions until after the problems were presented. Although interference from the decision task may explain subjects' poor performance on the recognition test, it is only a partial explanation for the ineffectiveness of the debiasing instructions.

Instructions to ignore baserates and rely on likelihood information were presented at the beginning of the experiment prior to attempting any of the problems. Such instructions may lack salience unless accompanied by an example problem or closely linked to the specific judgment task. Furthermore, because the five judgment problems appear to represent a heterogeneous set of tasks, different instructions may have to be constructed for personality and statistical problems (e.g., the Tom problem versus the Cab problem). However, the most significant limitation to

debiasing instructions is the nature of the decision task. Whether one conceptualizes responses to base-rate problems as arising from representative thinking, experimental demands, or some other source, instructions alone are probably insufficient to convey to subjects how the problem must be structured in order to disengage from common practices. Sharpe and Adair (In press) found that instructions to remediate the hindsight bias were effective only after a set of items had been completed following such instructions and demands to utilize such instructions were made explicit for a second set of problems.

Experiment 5

Manipulation of the problem frame in the fourth experiment did not have a significant effect on either responses to the problems nor subjects' interpretations of the judgment task. A more common method of arriving at a "better" judgment in the real-world is to discuss the problem with others. However, groups sometimes do not make better decisions than their individual members (e.g., groupthink; Janis, 1971) and may be susceptible to biases in the decision process (e.g., group polarization; Isenberg, 1986). Nonetheless, group interaction offers an opportunity for subjects to express and validate their interpretation of the experimental task, and although the "correct" interpretation is often the minority viewpoint, sometimes the minority opinion does prevail (Moscovici, 1985).

Only a handful of studies have examined the effects of group discussion on base-rate judgments, and the findings from these studies are mixed. Argote, Seabright and Dyer (1986) had subjects make probability judgments on three engineer/lawyer-type problems for which the individuating information was varied. Subjects worked either by themselves (coacting) or in groups of five. For stereotypic descriptions, groups relied slightly more on individuating information than did individuals, and thus the former were somewhat more likely to deviate from the normative solution.

Argote, Devadas and Melone (1990) included five-person

groups, coacting individuals, and individual protocol/no protocol problem solving groups. For stereotypic descriptions, individual judgments were superior to group judgments. However, the reverse was found for neutral and for non-stereotypic target descriptions. Protocol analysis suggested that stereotypic descriptions were associated with a greater number of statements that reflected the similarity between the target and the occupational class.

Along the same lines, Wright and Wells (1985) explored the effects of group discussion on the dispositional bias. Subjects were assigned to make an attitude attribution after working alone, working alone but spending 10 minutes thinking about their answer with or without anticipation of engaging in a discussion, or engaging in a discussion. Only for the last condition was there a significantly reduced dispositional bias. What specific aspects of the group discussion influenced judgments were not clear. However, the most obvious factor, reference to the choice/no choice component of the attitude attribution paradigm, was not found to mediate responses.

In the first of a series of experiments, Wright et al., (1990) determined group discussion increased use of consensus information. A subsequent study found that subjects who read arguments generated during the group discussion, but were not participants in the discussion, also utilized consensus information.

The four previous studies interpreted their findings from a persuasive arguments perspective. According to Argote and colleagues (Argote et al., 1986, 1990), individuals are more likely to become polarized toward the choice with a greater number of supporting arguments made during discussion. Because most individuals believe the individuating information is more useful, the group disregards the base-rate information. However, Wright and colleagues (Wright & Wells, 1985; Wright et al., 1990) noted that at least a few group members mentioned consensus information. Group members may have shifted toward the consensus information response as a function of the novelty and persuasiveness of that argument.

An obvious difference between the studies of Argote et al., and Wright et al., is that the former were investigating probability judgments, while the latter were examining social judgments. A less obvious difference between the two sets of studies is that Argote et al., required groups of subjects to give a single response. In contrast, Wright et al., had each member of the group make their own final judgment. The response of one member of the group then was randomly selected for inclusion in the data analysis.

Like alternative versions of the base-rate problems in Experiment 3 and instructions to modify the experimental frame in Experiment 4, group discussion has been found to

sometimes reduce deviations from the normative solution, but equally often has not been found to be effective. When group discussion has been found to reduce deviations from the normative solution, it is unclear what has mediated this effect. One possibility is that group discussion may result in subjects adopting an interpretation of the experimental task more consistent with the normative interpretation.

Half of the discussion groups in the present study were of the type in which members of the group were required to agree on a judgment. The other half asked group members to make their own judgments after engaging in discussion. It was hypothesized that group discussion should reduce the deviation from the normative solution for the Bret problem, an attitude attribution task, but increase the deviation from the normative solution for the Tom W. and engineer/lawyer problems, both representativeness judgment tasks. It was also predicted that deviations from the normative solution should be greater when groups of subjects were required to arrive at a single solution than when each member of the group was permitted to make his or her own judgment.

The present study also examined how subjects' interpretations of the judgment task might be modified by group discussion. The phenomenological correlates are essential because even when effects for group discussion have been found, the source of these effects have been

unclear. Subjects may be responding to the demands of the group, i.e., group polarization. Alternatively, subjects may modify their weighting of the evidential variables, i.e., seeing the base-rate information as more important. Finally, subjects may interpret the task differently after group discussion, i.e., discovering that the likelihood information can be ignored.

Method

Subjects. One-hundred and thirty-one first-year introductory psychology students from the University of Manitoba subject pool participated in Experiment 5. Subjects received credit for their research participation.

Materials and Procedure. Subjects were assigned to groups of 3 to 5 participants. One-third of the groups completed the problems without group discussion. The procedure from Experiment 1 was followed except that subjects were required to spend two minutes considering each problem before recording their response. All other subjects had the opportunity to spend two minutes discussing each problem before responding either on their own or as a group. All subjects then individually completed the PEQ from earlier studies.

Results

Responses to the Judgment Problems. Summary statistics for responses to the judgment problems by groups are presented in Table 23. To assess whether there were any

Table 23

Summary Statistics for the Five Judgment Problems by Instructions to Group, Experiment 5

	Cab	Jack	Linda	Tom	Bret
Group/No Consensus					
M_d	29.68	39.58	2.96	4.84	4.29
S_d	8.73	17.63	1.77	0.54	0.88
Conf	4.88	5.71	5.36	5.65	5.46
Group/Consensus					
M_d	35.00	40.00	4.11	5.22	4.81
S_d	10.00	24.49	1.17	0.66	0.99
Conf	5.39	5.08	4.96	6.21	5.67
Individual					
M_d	33.57	45.36	4.75	4.82	4.75
S_d	6.75	17.19	0.57	0.56	0.72
Conf	4.68	5.71	5.30	5.61	5.70
Combined					
M_d	32.72	42.22	4.07	4.92	4.63
S_d	8.21	18.84	1.38	0.59	0.84
Conf	4.90	5.57	5.24	5.76	5.62

Note. The number of groups were 12, 9 and 19 for Group with Consensus, Group Judgment without Consensus, and Individual Judgment, respectively.

effects for group discussion on base-rate judgments, a series of one-way ANOVAs were calculated on the absolute deviation scores, with group composition as the between-subject factor (Table 24). After calculation of the ANOVA, two planned comparisons were conducted. The first compared group discussion (with and without consensus) to individual judgments. The second compared group discussion with consensus to group discussion without consensus.

Only for the Linda problem was the overall ANOVA significant (Table 24). Tukey post-hoc tests indicated absolute deviation scores for the discussion without consensus group ($\underline{M}$ = 2.96) were significantly smaller than those of the individual judgment group ($\underline{M}$ = 4.75). The first contrast was also significant for the Linda problem. The second contrast was significant for Linda and marginally significant for Bret. For both problems, responses made after group discussion but without consensus were significantly less deviant from the normative solution compared to group judgment with consensus (Table 23).

A different method for constructing the dependent scores for the preceding analysis was to represent each group by one randomly selected subject rather than by the mean for the group. Employing the former procedure, there was no significant main effect for group composition for the Cab ($\underline{F}$ (1,38) = 0.32, $\underline{MS}_e$ = 171.85, $\underline{p}$ <.57), Jack ($\underline{F}$ (1,39)

Table 24

ANOVA Results and Planned Comparisons for Differences in the Absolute Deviation Scores for the Five Judgment Problems by Group Composition, Experiment 5.

Problem	df	MS_e	F	p
Cab				
Overall	2,37	66.45	1.29	.29
Comp 1	1,37		2.19	.15
Comp 2	1,37		0.22	.64
Jack				
Overall	2,37	365.85	0.43	.65
Comp 1	1,37		0.00	.97
Comp 2	1,37		0.84	.36
Linda				
Overall	2,37	1.38	8.55	.001
Comp 1	1,37		4.90	.03
Comp 2	1,37		10.61	.002
Tom				
Overall	2,37	0.33	1.60	.21
Comp 1	1,37		2.24	.18
Comp 2	1,37		1.26	.27

Bret				
Overall	2,37	0.69	1.42	.25
Comp 1	1,37		2.03	.16
Comp 2	1,37		0.58	.45

Note. Comp 1 is the planned comparison of discussion (with and without consensus) to individual judgment. Comp 2 is the planned comparison of group discussion with consensus to group discussion without consensus.

= 0.27, MS_e = .27, p <.61), Tom (F (1,39) = 0.93, MS_e = .93, p <.20) and Bret (F (1,39) = 0.00, MS_e = 1.44, p <.95) problems. The Linda problem was again significant (F (1,39) = 7.12, MS_e = 2.33, p < .01). The absolute deviation score was larger for individual judgments (M = 4.75) compared to group judgments (M = 3.46).

In order to assess whether variability, rather than group means, differed between treatment conditions, Brown-Forsythe tests of homogeneity of variance (Keppel, 1991) were calculated for both the three group (no discussion/discussion with consensus/discussion without consensus) and the two group cases (discussion/no discussion), and for both a mean score representing the group judgment and an individual randomly selected to represent the group. For the three group/mean score comparison, the only significant ANOVA was for the Linda problem (F (2,37) = 5.58, MS_e =.59, p <.008). Consistent with the analysis of the absolute deviation scores, the within-cell variance of the discussion without consensus group (S = 2.96) was smaller than the variance of the individual judgment group (S = 4.75). For the three group/randomly selected score comparison, there was a significant ANOVA for the Jack (F (2,38) = 3.31, MS_e = 196.75, p < .05) problem, with group discussion with consensus being more variable than the other two groups.

Confidence Ratings. A mixed-model ANOVA was used to

examine differences in confidence ratings for discussion groups compared to individual judgments on confidence ratings. The between-subject factor of instructions was not significant (F (2,37) = 0.04, MS_e = 1.46, p < .96). The within-subject effect for problem was significant (F (4,148) = 4.93, MS_e = 0.81, p < .001). Tukey HSD post-hoc tests (p < .05) revealed that subjects were less confident in their response to the Cab problem (M = 4.9) than in their responses to the Jack (M = 5.57), Tom (M = 5.76) and Bret problems (M = 5.62).

Examining the confidence ratings from one subject randomly selected to represent each group, the between-subject effect for group composition was marginally nonsignificant (F (1,38) = 3.56, MS_e = 2.26, p < .07), but the within-subject effect for problem (F (4,152) = 3.07, MS_e = 1.75, p < .02) and the interaction of group composition and problem (F (4,152) = 2.57, MS_e = 2.57, p < .04) were significant. For both the Cab and Tom problems, subjects who engaged in group discussion were significantly more confident in their responses (Ms = 5.38 and 5.81) than subjects who had individually given their response (Ms = 4.21 and 4.84).

One ancillary analysis provided interesting results. Approximately half of the Group/No Consensus and Group/Consensus groups were composed of three members, the other half of four members. An analysis of variance was

conducted on absolute deviation scores and confidence ratings with Consensus/No Consensus and size of group serving as between-subject factors. There were no significant effects associated with the absolute deviation scores. For confidence ratings, however, there was a significant main effect for group size for the Jack (F (1,17) = 5.52, MS_e = .54, p < .03), and Bret problems (F (1,17) = 6.26, MS_e = .71, p < .02), a marginally nonsignificant main effect for group size for the Linda problem (F (1,17) = 2.91, MS_e = 1.09, p < .11), and a significant Consensus/No Consensus by group size interaction for the Cab (F (1,17) = 7.04, MS_e = 1.41, p < .02) and Jack problems (F (1,17) = 5.72, MS_e = .54, p < .03. Cell means are presented in Table 25. Groups of four participants were more confident in their judgments. The significant interactions for the Cab and Jack problems resulted from the higher confidence reported by Consensus subjects in four-person groups.

Solutions Demanded. Group discussion and individual judgments were compared for subjects' perceptions of whether there was a single correct response or more than one correct response, and whether an intuitive and personal judgment or a formal analysis was demanded. Group composition was a between-subject factor, problem was a within-subject factor. For number of correct responses, neither group composition (F (1,38) = 1.02, MS_e = 2.72, p <.32) nor the interaction of

Table 25

Mean Confidence Ratings for Each Judgment Problem by Group Size and Consensus/No Consensus, Experiment 5

	Problem				
Condition	Cab	Jack	Linda	Tom	Bret
No Consensus					
3-Person (n=6)	5.39	5.72	5.05	5.50	5.05
4-Person (n=6)	4.38	5.71	5.67	5.79	5.87
Consensus					
3-Person (n=5)	4.60	4.40	4.53	5.93	5.20
4-Person (n=4)	6.37	5.94	5.50	6.65	6.25

Note. Higher numbers reflect greater confidence. ns = number of groups. Five-person groups were excluded from this analysis.

group composition by problem (F (4,152) = 1.65, MS_e = 1.18, p <.16) were significant. Problem was a significant within-subject factor (F (4,152) = 9.16, MS_e = 1.18, p <.0001). Post-hoc tests revealed that the Cab problem (M = 2.94) was more demanding of a single correct solution compared to the Linda (M = 4.23) and Tom problems (M = 3.86), and the Jack (M = 3.39) and Bret problems (M = 3.32) were more demanding of a single correct response compared to the Linda problem.

Evaluating whether an intuitive judgment or a formal analysis was demanded, there was no significant main effect for group composition (F (1,38) = 0.40, MS_e = 1.59, p <.40) or for the interaction of group composition and problem (F (4,152) = 0.93, MS_e = 1.19, p <.45). There was a significant within-subject effect for problem (F (4,152) = 57.80, MS_e = 1.19, p <.0001). Subjects perceived the Cab problem (M = 5.77) to be significantly more demanding of a formal analysis than the Jack (M = 4.01), Linda (M = 2.61), Tom (M = 2.77) and Bret problems (M = 3.07). In turn, the Jack problem was seen to be more demanding of a formal analysis than the Linda, Tom and Bret problems.

Relevance of Statements. A series of mixed ANOVAs were employed to examine the effects of group composition on subjects perceptions of the relevance of statements derived from each problem. The only significant group composition by statement interaction was for the Cab problem (F (3,114) = 4.23, MS_e = 1.30, p < .007). The base-rate statement was

seen as significantly more relevant by subjects who made their judgment as individuals ($\underline{M}$ = 4.28) compared to subjects who made their judgment as a group ($\underline{M}$ = 3.6).

Discussion

Engaging in group discussion had little impact on performance with the exception of the Linda problem. For the Linda problem, group discussion resulted in smaller absolute deviations from the normative solution compared to individual judgments, but only when group consensus was not imposed. No specific prediction had been made for the effects of group discussion for the Linda problem. The Linda problem, more than any other problem, evokes a knew-it-all-along effect for at least some portion of subjects. Furthermore, it is probably easiest to defend the normative response to the Linda problem.

One limitation to ad-hoc group discussion is that unless such groups are explicitly (and continuously) prompted to engage in discussion, deliberation tends to be minimal. Larger groups offer a greater probability that one member will recognize that base-rate elements are important, and likelihood information should be ignored. On the other hand, some individuals are uncomfortable with speaking in small groups of strangers, and this tendency is likely to inhibit presentation of an argument contrary to the position of most members of the group. Group decisions were also found to be associated with greater confidence, and the

first showed that greater confidence was associated with deviations from the normative response. Confidence was also greater for members of four person groups over three person groups, although group size was not associated with smaller deviations from the normative response.

General Discussion

The most widely accepted explanation for subjects' poor performance on base-rate problems relates to how humans process complex information. Humans are portrayed as limited information processors who expend only the minimal amount of cognitive effort necessary to attain some satisfactory response. In the same vein, heuristics are thought to be employed over more complex processes because these mental short-cuts require minimal cognitive effort and usually result in a reasonable solution to many real-world problems (Tversky & Kahneman, 1974).

An alternative to the heuristic explanation explored in this series of experiments was that some subjects deliberately ignored base-rate information for at least some of the problems. This was attributed to subjects seeing baserates as irrelevant to achieving satisfactory (albeit non-normative) solutions to the problems. This explanation was most strongly supported by the results from problems that presented likelihood information as personality descriptions. For the Jack and Tom problems, base-rate statements were judged to be less relevant than likelihood statements, and the latter correlated negatively with deviation scores. For the Cab and Bret problems, however, base-rate and likelihood information were seen as equally relevant, and neither consistently correlated with deviation scores. These results also offer at least partial support

for the proposition that likelihood information is considered in social domains, and the base-rate is considered in non-social domains (Jacobs & Potzner, 1991).

Evidence suggesting subjects do not interpret judgment problems in the manner intended by experimenters was obtained by asking subjects to endorse one of two interpretations for each problem after giving their response (Experiment 3). Subjects endorsed the normative interpretations for the Cab, Linda and Bret problems, although their actual responses to these problems deviated from the respective normative solutions. This appears to represent what Kahneman and Tversky (1982) referred to as errors of application or failing to apply the correct procedure to solve a problem. In contrast, subjects endorsed non-normative interpretations for the solutions to the Jack and Tom problems, the two problems that most directly involve personality descriptions. Subjects were either unaware of the correct procedure or did not regard the correct procedure as appropriate.

Regardless of whether subjects made errors of comprehension or application, their endorsement and relevance ratings were calibrated with their solutions to the problems. The definition of rationality most widely accepted within the psychological literature is that a decision is rational if it does not violate the dictates of normative models (Dawes, 1988), such as probability theory

(e.g., Bayes' theorem). It follows from this definition that many of the subjects in these experiments may be regarded as irrational decision makers. However, the solutions subjects offered to the base-rate problems were consistent with their phenomenological reports. Thus, subjects may have been irrational decision makers in the probabilistic sense, but their experimental behavior was rational in that it followed from their reasoning as to the nature of the experimental task.

Less direct evidence for subjects' viewing the experimental task differently than intended was provided by the nature of solution items and by subjects' written protocols. Deviations from the normative solutions were associated with endorsement of intuitive rather than formal analyses of the problems. Only for the Cab problem was the normative interpretation of a single correct response and formal analysis found. Analysis of subjects' statements regarding the purpose of the study and rationales for high and low responses to the problems revealed subjects saw their task as involving person perception and careful examination of the likelihood information.

No direct effort was made to ascertain what was the exact nature of subjects' non-normative interpretations of the experimental task. However, the context of the task is a psychology experiment. Subjects are told psychologists are good at the task. They are presented with a vivid,

detailed, and one-sided personality description, and asked to judge whether some target individual belongs to some social group. Under those circumstances, it is appropriate to see likelihood information as more relevant than base-rate information (Schwarz et al., 1991).

Other findings from these studies were consistent with a model that attributes deviations from normative solutions to a non-normative interpretation of the experimental task. If situational factors within the experimental setting (and possibly factors specific to each problem) determine the interpretation adopted for the experimental task, then individual difference variables may be associated with the particular interpretation adopted, but not be directly related to deviation scores.

Confidence ratings were a dependent measure that were found to directly relate to subjects' responses to the base-rate problems. The more responses deviated from the normative solution, the more confident were subjects in those responses. Statistically sophisticated subjects, males, and decision making groups with four rather than three members, were also found to be more confident in their responses but not more accurate. If the "correct" answer to a judgment problem is obvious (Politzer & Noveck, 1991), then subjects <u>should</u> be more confident. On the other hand, if subjects are "suspicious" of the demands of the experimental task, and give responses inconsistent with

these demands but correct according to the normative model, then they should be less confident in their responses.

Several approaches to reducing deviation scores were attempted in order to facilitate an understanding of why deviations from the normative solution arise. Consistent with the results from many previous studies, these attempts met with mixed success. For example, telling subjects to ignore likelihood and to use base-rate information (Experiment 4) likely contradicted what may have been a powerful experimental message to use all of the information provided. Subjects also did not pay close attention to these instructions. This may have been because of their placement at the beginning of the experimental booklet where instructions could lack meaning prior to attempting the problems.

Group discussion was also not found to affect the magnitude of deviation scores (Experiment 5). For group discussion to have been effective, at least one member of the discussion group must have possessed an interpretation of the experimental task consistent with the normative model, and been willing and able to convince other members of the group of the validity of this interpretation. Anyone who has attempted to convince naive problem solvers of the virtues of the normative interpretation will attest to the difficulty of this undertaking.

The only successful approach to reducing deviations

from the normative response was to alter the likelihood information (Experiment 3). This is similar to a debiasing procedure where outcome feedback is discredited by saying an error was made by the experimenter (Hasher, Attig & Alba, 1979). This approach to reducing bias does not contradict the experimental message to use likelihood information, but rather affects the diagnosticity or value of that information (Hilton & Fein, 1989).

Do subjects sometimes examine the similarity between the characteristics of an individual and a social group as posited by the representativeness thesis? There is much anecdotal and some empirical evidence to suggest that subjects in the real-world often make decisions on this basis (Nisbett, Krantz, Jepson, & Kunda, 1983). Examination of subjects' written rationales for high and low responses to the laboratory problems also revealed many instances of representative thinking (Experiment 3). The representativeness heuristic may be invoked in laboratory demonstrations by experimental demands (Krosnick, et al., 1990) to use the likelihood and ignore the base-rate information rather than arising spontaneously from the cognitive demands of the decision task. However, it does not follow that laboratory demonstrations of base-rate neglect are mere "laboratory curiosities" (Kahneman, 1991) unrelated to real-world decision-making. The experimental laboratory is but one of many social contexts, although it

may make unique demands upon participants, and may more clearly differentiate base-rate from likelihood information (c.f., Wells, 1992).

Responses across the five problems were not found to correlate significantly with each other (Experiment 1). This between-problem variability implies subjects may use multiple strategies to solve a set of base-rate problems, and suggests care be taken in speaking of a general base-rate fallacy. The same issue has arisen in other problem solving contexts. Like base-rate problems, insight problems are said to demand subjects gain some understanding of their underlying structure (Metcalfe & Wiebe, 1987). Metcalfe (1986) found subjects to be unable to accurately predict whether or not they would correctly solve insight problems. However, Sharpe and Ritchot (1991) found some insight problems to be associated with strong feeling-of-knowing, and attributed this contradiction to previous studies collapsing across dissimilar problems.

Consideration should be given to several factors that limit the conclusions drawn from this series of studies. Bayes' theorem was the standard by which subjects' judgments were evaluated. Conclusions drawn from the data would depend, of course, on the suitability of Bayes' theorem as the normative standard. A second limitation is the artificiality of the decision problems and the reliance on introductory students as subjects. Real-world decision

tasks and individuals with greater experience at decision making would produce different results. A third and more general limitation is that this dissertation was undertaken within a logical positivist tradition that has been criticized for having failed to consider the meaning subjects give to their experimental behavior. Limitations to that research approach must apply to this dissertation.

One recent trend in the judgmental bias literature has been to downplay the cataloguing of biases in human judgment, such as the base-rate fallacy, but rather to devote effort to explaining how decision making is accomplished (Hastie, 1991). Another trend has been to try to incorporate non-cognitive factors into decision making (Kahneman, 1991), such as the social context in which decision making takes place. Most directly relevant to the research reported in this set of experiments, however, is the trend to consider how conversational logic (Schwarz, et al., 1991), rules (Politzer & Noveck, 1991), conventions (Krosnick et al., 1990), and processes (Hilton, 1990) affect judgment. It is important to explore how subjects construe and react to the experimental situation and the decision task. According to Hastie (1991), "the most important direction for research in the next few decades is to describe the manner in which people 'make sense' of their situation as a preliminary to making a decision" (p. 137). If this research direction is taken, the perspectives of subjects and experimenters should converge, and a better understanding of the decision process should result.

References

Adair, J. G. (1982). Meaning of the situation to subjects. American Psychologist, 37, 1406-1408.

Adair, J. G. (1991). Social cognition, artifact and the passing of the so-called crisis in social psychology. Canadian Psychology, 32, 445-450.

Adair, J. G., & Fenton, D. P. (1971). Subject's attitudes toward psychology as a determinant of experimental results. Canadian Journal of Behavioral Science, 3, 268-275.

Adler, J. E. (1984). Abstraction is uncooperative. Journal for the Theory of Social Behavior, 14, 165-180.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. Cognitive Psychology, 21, 515-550.

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

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. Psychological Bulletin, 82, 261-277.

Allwood, C. M., & Bjorhag, C. G. (1990). Are two judges better than one? On the realism in confidence judgements by pairs and individuals. In J. P. Caverni, J. M. Fabre & M. Gonzalez (Eds.), Cognitive biases (pp. 443-463). New York: North Holland.

Argote, L., Devadas, R., & Melone, N. (1990). The base-rate fallacy: Contrasting processes and outcomes of group and individual judgments. Organizational Behavior and Human Decision Processes, 46, 296-310.

Argote, L., Seabright, M. A., & Dyer, L. (1986). Individual versus group use of base-rate and individuating information. Organizational Behavior and Human Decision Processes, 38, 65-75.

Asch, S. (1952). Social psychology. Englewood Cliffs: Prentice Hall.

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

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

Bar-Hillel, M., & Fischhoff, B. (1981). When do base rates affect predictions? Journal of Personality and Social Psychology, 41, 671-680.

Barnes, V. E. (1985). The quality of human judgment: An alternative perspective. Unpublished doctoral dissertation, University of Washington.

Beach, L. R., Christensen-Szalanski, J., & Barnes, V. (1987). Assessing human judgment: Has it been done, can it be done, should it be done? In G. Wright & P. Ayton (Eds.), Judgmental forecasting (pp. 49-62). New York: Wiley.

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

Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. American Psychologist, 37, 245-257.

Berkowitz, L., & Troccoli, B. T. (1986). An examination of the assumptions in the demand characteristics thesis: With special reference to the Velten Mood Induction Procedure. Motivation and Emotion, 10, 337-349.

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

Borgida, E., & Brekke, N. (1981). The base-rate fallacy in attribution and prediction. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), New directions in attribution research (Vol. 3, pp. 63-95). Hillsdale, NJ: Erlbaum.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). A study of thinking. New York: Wiley.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. Journal of Personality and Social Psychology, 42, 116-131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. Journal of Personality Assessment, 48, 306-307.

Campbell, D. (1966). Factors relevant to the validity of experiments in social settings. In C. W. Backman & P. F. Secord (Eds.), Problems in social psychology (pp. 3-13). New York: McGraw-Hill.

Cantor, N., Pittman, T. S., & Jones, E. E. (1982). Choice and attitude attributions: The influence of constraint information on attributions across levels of generality. Social Cognition, 1, 1-20.

Carlson, B. W., & Yates, J. F. (1989). Disjunction errors in qualitative likelihood judgment. Organizational Behavior and Human Decision Processes, 44, 368-379.

Carroll, J. S., & Siegler, R. S. (1977). Strategies for the use of base-rate information. Organizational Behavior and Human Performance, 19, 392-402.

Cohen, J. (1981). Can human irrationality be experimentally demonstrated? Behavior and Brain Sciences, 4, 317-370.

Cohen, 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, 250-274.

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

Devine, P. G. (1989). Overattribution effect: The role of confidence and attributional complexity. Social Psychology Quarterly, 52, 149-158.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. Journal of Personality and Social Psychology, 58, 568-581.

Ebbesen, E. B., & Konecni, V. J. (1980). On the external validity of decision-making research: What do we know about decisions in the real world? In T. S. Wallsten (Ed.), Cognitive processes in choice and decision behavior (pp. 21-45). Hillsdale, NJ: Erlbaum.

Edwards, W. (1982). A view from a barefoot decision analyst. In G. R. Ungson & D. N. Baunstein (Eds.), Decision making: An interdisciplinary inquiry (pp. 317-320). Boston: Kent Publishing.

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., & Pollard, P. (1982). Statistical judgment: A further test of the representativeness construct. Acta Psychologica, 51, 91-103.

Farrington, D. P. (1980). External validity: A problem for social psychology. In R. F. Kidd & M. J. Saks (Eds.), Advances in Applied Social Psychology (pp. 184-186). Hillsdale, NJ: Erlbaum.

Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motivation and the correspondence bias. Journal of Personality and Social Psychology, 58, 753-764.

Fein, S., & Hilton, J. L. (1992). Attitudes toward groups and behavioral intentions toward individual group members: The impact of nondiagnostic information. Journal of Experimental Social Psychology, 28, 101-124.

Fischhoff, B. (1977). Perceived informativeness of facts. Journal of Experimental Psychology, 3, 349-358.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 422-444). New York: Cambridge.

Fischhoff, B. (1990). Discussion. In R. M. Hogarth (Ed.), Insights in decision making: A tribute to Hillel J. Einhorn (pp. 337-342). Chicago: University of Chicago Press.

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

Fischhoff, B., & Bar-Hillel, M. (1984b). Focusing techniques: A shortcut to improving probability judgments? Organizational Behavior and Human Performance, 34, 175-194.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. Psychological Review, 90, 239-260.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perceptions and Performance, 3, 552-564.

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

Fiske, S. T., & Taylor, S. E. (1984). Social cognition. Reading MA: Addison-Wesley.

Fletcher, G. J. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual difference measure. Journal of Personality and Social Psychology, 51, 875-884.

Fletcher, G. J., Reeder, G. D., & Bull, V. (1990). Bias and accuracy in attitude attribution: The role of attributional complexity. Journal of Experimental Social Psychology, 26, 275-288.

Frisch, D. E. (1988). Violations of probability theory: What do they mean? Journal for the Theory of Social Behaviour, 18, 137-148.

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

Furnham, A., Jaspars, J., & Fincham, F. D. (1983). Professional and naive psychology: Two approaches to the explanation of social behaviour. In J. Jaspars, F. D. Fincham & M. Hewstone (Eds.), Attribution theory and research: Conceptual, developmental and social dimensions (pp. 315-340). New York: Academic.

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

Gigerenzer, G. (1991b). How to make cognitive illusions disappear: Beyond "heuristics and biases". In W. Stroebe & M. Hewstone (Eds.), European Review of Social Psychology (Vol 2., pp. 83-115). New York: John Wiley.

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

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

Gigerenzer, G., Swijtink Z., Porter T., Daston L., Beatty, J. & Kruger, L. (1989). The empire of chance: How probability changed science and everyday life. Cambridge: Cambridge University Press.

Ginosar, Z., & Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. Journal of Experimental Social Psychology, 16, 228-242.

Gravetter, F. J., & Wallnau, L. B. (1988). Statistics for the behavioral sciences: A first course for students of psychology and education (2nd ed.). St. Paul: West.

Greenberg, J., & Folger, R. (1988). Controversial issues in social research methods. New York: Springer-Verlag.

Grice, H. P. (1975). Logic and conversation. In D. Davidson & G. Harman (Eds.), The logic of grammar (pp. 64-74). Encino, CA: Dickenson.

Harre, R., & Secord, P. F. (1972). The explanation for social behaviour. Oxford: Basil Blackwell.

Hasher, L., Attig, M. S., & Alba, J. W. (1979). I knew it all along: Or, did I? Journal of Verbal Learning and Verbal Behavior, 20, 86-91.

Hastie, R. (1991). A review from a high place: The field of judgment and decision making as revealed in its current textbooks. Psychological Science, 2, 135-138.

Hastie, R., Penrod, S. D., & Pennnington, N. (1983). Inside the jury. Cambridge, MA: Harvard University Press.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Hicks, L. E. (1985). Is there a disposition to avoid the fundamental attribution error? Journal of Research in Personality, 19, 436-456.

Hilton, D. J. (1990). Conversational processes and causal explanation. Psychological Bulletin, 107, 65-81.

Hilton, J. L., & Fein, S. (1989). The role of typical diagnosticity in stereotype-based judgments. Journal of Personality and Social Psychology, 57, 201-211.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. Psychological Review, 93, 75-88.

Hinsz, V. B., Tindale, R. S., Nagao, D. H., Davis, J. H., & Robertson, B. A. (1988). The influence of the accuracy of individuating information on the use of base rate information in probability judgments. Journal of Experimental Social Psychology, 24, 127-145.

Hogarth, R. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. Psychological Bulletin, 90, 197-213.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. Journal of Personality and Social Psychology, 50, 1141-1151.

Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. Child Development, 62, 166-178.

Janis, I. L. (1971, November). Groupthink. Psychology Today, pp. 43-46.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. Journal of Experimental Social Psychology, 3, 1-24.

Jones, E. E., Riggs, J. M., & Quattrone, G. (1979). Observer bias in the attitude attribution paradigm: The effect of time and information order. Journal of Personality and Social Psychology, 37, 1230-1238.

Jones, E. E., Worchel, S., Goethals, G. R., & Grumet, J. F. (1971). Prior expectancy and behavioral extremity as determinants of attitude attribution. Journal of Experimental Social Psychology, 37, 1230-1238.

Jungermann, H. (1986). The two camps on rationality. In H. R. Arkes & K. R. Hammond (Eds.), Judgment and decision making: An interdisciplinary reader (pp. 627-641). Cambridge: Cambridge University Press.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D. (1991). Judgment and decision making: A personal view. Psychological Science, 2, 142-145.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgments under uncertainty: Heuristics and biases (pp. 493-508). Cambridge: Cambridge University Press.

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

Karlsson, G. (1989). Rules and strategies in decision making: A critical analysis from a phenomenological perspective. In H. Montgomery & O. Svenson (Eds.), Process and structure in human decision making (pp. 51-64). New York: Wiley.

Kassin, S. M. (1979). Consensus information, prediction, and causal attribution: A review of the literature and issues. Journal of Personality and Social Psychology, 37, 1966-1981.

Kelley, H. H. (1972). Attributions in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior (pp. 1-26). Morristown, N.J.: General Learning Press.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Konold, C. (1989). Informal conceptions of probability. Cognition and Instruction, 6, 59-98.

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 and Social Psychology, 59, 1140-1152.

Krosnick, J. A., & Sedikides, C. (1990). Self-monitoring and self-protective biases in use of consensus information to predict one's own behavior. Journal of Personality and Social Psychology, 58, 718-728.

Kruglanski, A. W., & Ajzen, I. (1983). Bias and error in human judgment. European Journal of Social Psychology, 19, 448-468.

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

Langer, E. J. (1989). Minding matters: The consequences of mindlessness-mindfulness. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 22, pp. 137-173). New York: Academic Press.

Langer, E., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. Journal of Personality and Social Psychology, 36, 635-642.

Leddo, J., Abelson, R. P., & Gross, P. H. (1984). Conjunctive explanations: When two reasons are better than one. Journal of Personality and Social Psychology, 47, 933-943.

Levine, D. W., & Dunlap, W. P. (1982). Power of the F test with skewed data: Should one transform or not? Psychological Bulletin, 92, 272-280.

Lindeman, S. T., Van den Brink, W. P., & Hoogstraten, J. (1988). Effects of feedback on base-rate utilization. Perceptual and Motor Skills, 67, 343-350.

Lynch, J. G., & Ofir, C. (1989). Effects of cue consistency and value on base-rate utilization. Journal of Personality and Social Psychology, 56, 170-181.

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

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

Meehl, P. E. (1957). When shall we use our heads instead of the formula? Journal of Counseling Psychology, 4, 268-273.

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 288-294.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. Memory and Cognition, 15, 238-246.

Miller, A. G. (1974). Perceived freedom and the attribution of attitudes. Representative Research in Social Psychology, 5, 61-79.

Miller, A. G., & Roser, L. G. (1982). Toward an understanding of the fundamental attribution error: Error diagnosticity in the attitude attribution paradigm. Journal of Research in Personality, 16, 41-59.

Miller, A. G., Schmidt, D., Meyer, C., & Colella, A. (1984). The perceived value of constrained behavior: Pressures toward biased inference in the attitude attribution paradigm. Social Psychology Quarterly, 47, 160-171.

Montgomery, H., & Svenson, O. (1989). Process and structure in human decision making. New York: Wiley.

Morier, D. M., & Borgida, E. (1984). The conjunction fallacy: A task specific phenomenon? Personality and Social Psychology Bulletin, 10, 243-252.

Moscovici, S. (1985). Social influence and conformity. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology, (3rd ed., pp. 347-412). Hillsdale, N.J.: Erlbaum.

Nelson, T. E., Biernat, M. R., & Manis, M. (1990). Everyday base rates (sex stereotypes): Potent and resilient. Journal of Personality and Social Psychology, 59, 664-675.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 90, 339-363.

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

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 17, 776-783.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), Artifact in behavioral research (pp. 143-179). New York: Academic Press.

Orne, M. T. (1973). Communication by the total experimental situation: Why is it important, how it is evaluated, and its significance for the ecological validity of findings. In P. Pliner, L. Krames, & T. Alloway (Eds.), Communication and affect (pp. 157-191). New York: Academic Press.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. Psychological Bulletin, 68, 29-46.

Pitz, G. F. (1980). The very guide of life: The use of probabilistic information for making decisions. In T. S. Wallsten (Ed.), Cognitive processes in choice and decision behavior (pp. 77-94). Hillsdale, NJ: Erlbaum.

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

Pollard, P., & Evans, J. St. B. T. (1983). The role of 'representativeness' in statistical inference: A critical appraisal. In J. St. B. T. Evans (Ed.), Thinking and reasoning: Psychological approach (pp. 107-134). London: Routledge & Kegan Paul.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. Journal of Personality and Social Psychology, 42, 593-607.

Reeves, T. L., & Lockhart, R. S. (1990, June). The disjunction fallacy in probabilistic reasoning. Paper presented at the annual convention of the Canadian Psychological Association, Ottawa, Canada.

Roberts, L. D. (1991). Relevance as an explanation of communication. Linguistics and Philosophy, 14, 453-472.

Ross, L. (1978). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Cognitive theories in social psychology (pp. 337-383). New York: Academic Press.

Scheidt, R. J. (1981). Ecologically-valid inquiry: Fait accompli? Human Development, 24, 225-228.

Schneider, D. J., & Miller, R. S. (1975). The effects of enthusiam (sic) and quality of arguments on attitude attribution. Journal of Personality, 43, 693-708.

Scholz, R. W. (1987). Cognitive strategies in stochastic thinking. Dordrecht: D. Reidel.

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.

Sharpe, D., & Adair, J. G. (In press). Reversibility of the hindsight bias: Manipulation of experimental demands. Organizational Behaviour and Human Decision Processes.

Sharpe, D., & Ritchot, K. (1991, June). Feeling-of-knowing and knew-it-all-along effects in insight problem solving. Paper presented at the annual meeting of the Canadian Psychological Association, Calgary, Canada.

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer & T. K. Srull (Eds.), Handbook of social cognition (Vol. 1, pp. 189-286). New York: Erlbaum.

Snyder, M. (1974). Self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 30, 526-537.

Snyder, M., & Jones, E. E. (1974). Attitude attribution when behavior is constrained. Journal of Experimental Social Psychology, 10, 585-600.

Sperber, D., & Wilson, D. (1986). Relevance. Cambridge: Basil Blackwell.

Svenson, O. (1989). Eliciting and analysing verbal protocols in process studies of judgment and decision making. In H. Montgomery & O. Svenson (Eds.), Process and structure in human decision making (pp. 65-82). New York: Wiley.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics (2nd ed). New York: Harper and Row..

Treisman, A. M. (1969). Strategies and models of selective attention. Psychological Review, 76, 282-299.

Turnbull, W., & Slugoski, B. R. (1988). Conversational and linguistic processes in causal attribution. In D. J. Hilton (Ed.), Contemporary science and natural explanation: Commonsense conceptions of causality (pp. 66-93). New York: New York University Press.

Tversky A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1130.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Tversky, A., & Kahneman, D. (1982a). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgments under uncertainty: Heuristics and biases (pp. 153-162). Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982b). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgments under uncertainty: Heuristics and biases (pp. 84-98). Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.

von Neumann, J., & Morgenstern, O. (1953). Theory of games and economic behavior. (2nd ed.), New York: Wiley.

Wells, G. L., & Harvey, J. H. (1978). Naive attributors' attributions and predictions: What is informative and when is an effect an effect? Journal of Personality and Social Psychology, 36, 483-490.

Wimsatt, W. C. (1986). Heuristics and the study of human behavior. In D. W. Fiske & R. A. Shweder (Eds.), Metatheory in social science (pp. 293-314). Chicago: University of Chicago Press.

Wright, E. F., Luus, E., & Christie, S. (1990). Does group discussion facilitate the use of consensus information in making causal attributions? Journal of Personality and Social Psychology, 59, 261-269.

Wright, E. F., & Wells, G. L. (1985). Does group discussion attenuate the dispositional bias? Journal of Applied Social Psychology, 15, 531-546.

Wright, E. F., & Wells, G. L. (1988). Is the attitude-attribution paradigm suitable for investigating the dispositional bias? Personality and Social Psychology Bulletin, 14, 183-190.

Wright, G., & Ayton, P. (1990). Biases in probabilistic judgment: A historical perspective. In J. P. Caverni, J. M. Fabre & M. Gonzalez (Eds.), Cognitive biases (pp. 425-442). New York: North Holland.

Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including "signed summation". Organizational Behavior and Human Decision Processes, 37, 230-253.

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

APPENDIX A

Age _____ Sex _____

This booklet contains a number of problems to be solved. The answers to the problems should be recorded in this booklet in the spaces provided.

Please go through the booklet page by page. **Do not go on to the next page without completing the current page and do not return to previous pages.**

Take care in approaching each problem. Your score will be reflected in how close you come to the correct answer. Work alone and please do not ask any questions until all booklets are completed.

A cab was involved in a hit-and-run accident at night. Two companies, the Green and the Blue, operate in the city. You are given the following data:

(i) 85% of the cabs in the city are Green and 15% are Blue.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

Question: What is the probability (expressed as a percentage) that the cab involved in the accident was Blue rather than Green?

_____%

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. From the descriptions, one man has been randomly selected and is described below.

> Jack is a 45 year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

The probability that Jack is one of the 30 engineers in the sample of 100 is

_____%.

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

Linda is a 31 year old, single, outspoken, and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.

Is it more probable that Linda is a

1	2	3	4	5	6	7
Bank Teller						Bank Teller and a Feminist

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.

Is it more probable that Tom is now a graduate student in

1	2	3	4	5	6	7
Education						Library Science

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

Bret was a student in social psychology last year. Brief essays were written on the topic of whether there should be more or less funding for research into AIDS. Bret was assigned to write a brief essay arguing for less funding for research into AIDS. This is his essay.

> As usual when a new threatening disease is discovered it attracts more media attention than it deserves. Acquired Immune Deficiency Syndrome (AIDS) is comparatively rare and did not warrant the coverage in the media it has received. An enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases. If that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.

Bret is _____ increasing funding for research into AIDS.

1	2	3	4	5	6	7
Strongly Against						Strongly For

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

Perceptions of the Experimental Task

The first part of the experiment is now over. However, the following questions are an important part of this study. They are designed to help me understand your responses to the first part of the study. One of my concerns is to attempt to understand how you perceived each of the problems I presented to you. There are no right or wrong answers to the questions that follow. I am interested in your views of the problems during the experiment.

I don't want you to analyze the problems again. I want you to think back to when you first read the problems. What did you think of the problems, what did you attend to, and what did you base your answers upon? On the next few pages, please tell me honestly what thoughts you had about the experiment.

What do you think was the purpose of this experiment?

_____

_____

_____

Why did you think I asked you to solve these problems?

_____

_____

_____

What quality or skill of yours do you think I was trying to assess with these problems?

_____

_____

_____

Did you feel that the problems really didn't have a correct response? If so, which problems and why?

_____

_____

_____

Did you think there was any trick to answering the problems?

_____

_____

_____

Do you think there was a single correct response or more than one correct response to the cab problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the cab problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

A cab was involved in a hit-and-run accident at night. Two companies, the Green and the Blue, operate in the city. You are given the following data:

(i) 85% of the cabs in the city are Green and 15% are Blue.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

Question: What is the probability (expressed as a percentage) that the cab involved in the accident was Blue rather than Green?

_____%.

Please rate each of the following statements as to their relevance to your arriving at a solution to the cab problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ A cab was involved in a hit-and-run accident at night.

_____ 85% of the cabs in the city are Green and 15% are blue.

_____ A witness identified the cab as a Blue cab.

_____ The witness made correct identifications in 80% of the cases and erred in 20% of the cases.

Do you think there was a single correct response or more than one correct response to the engineer/lawyer problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the engineer/lawyer problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. From the descriptions, one man has been randomly selected and is described below.

> **Jack is a 45 year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.**

The probability that Jack is one of the 30 engineers in the sample of 100 is

_____%.

Please rate each of the following statements as to their relevance to your arriving at a solution to the engineer/lawyer problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ A panel of psychologists have interviewed and administered personality tests and prepared thumbnail sketches.

_____ 30 engineers and 70 lawyers, all successful

_____ Jack is a 45 year-old man, married with four children.

_____ He is generally conservative, careful and ambitious.

_____ He shows no interest in political and social issues.

_____ He spends most of his free time on his many hobbies: home carpentry, sailing, and mathematical puzzles.

Do you think there was a single correct response or more than one correct response to the Linda problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the Linda problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

Linda is a 31 year old, single, outspoken, and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.

Is it more probable that Linda is a

1 2 3 4 5 6 7

Bank Teller — **Bank Teller** and a Feminist

Please rate each of the following statements as to their relevance to your arriving at a solution to the Linda problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ Linda is a 31 year-old, single, outspoken, and very bright. She majored in philosophy.

_____ As a student she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.

Do you think there was a single correct response or more than one correct response to the Tom W. problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the Tom W. problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

> **Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.**

Is it more probable that Tom is now a graduate student in

1 2 3 4 5 6 7

Education — Library Science

Please rate each of the following statements as to their relevance to your arriving at a solution to the Tom W. problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

_____ Tom W. is of high intelligence, although lacking in true creativity.

_____ He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place.

_____ His writing is rather dull and mechanical, occassionally enlivened by somewhat corny puns and flashes of imagination.

_____ He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others.

_____ Self-centered, he nonetheless has a deep moral sense.

Do you think there was a single correct response or more than one correct response to the Bret problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the Bret problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

Bret was a student in social psychology last year. Brief essays were written on the topic of whether there should be more or less funding for research into AIDS. Bret was assigned to write a brief essay arguing for less funding for research into AIDS. This is his essay.

> **As usual when a new threatening disease is discovered it attracts more media attention than it deserves. Acquired Immune Deficiency Syndrome (AIDS) is comparatively rare and did not warrant the coverage in the media it has received. An enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases. If that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.**

Bret is _____ increasing funding for research into AIDS.

1 2 3 4 5 6 7

Strongly Against — Strongly For

Please rate each of the following statements as to their relevance to your arriving at a solution to the Bret problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ Bret was assigned to write a brief essay arguing for less funding for research into AIDS.

_____ Bret wrote that AIDS is comparatively rare and did not warrant the coverage in the media it has received.

_____ Bret wrote that an enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases.

_____ Bret wrote that if that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.

Have you seen any of the problems before?

_____

If so, which problems and where did you see them before?

_____

_____

Do you have any other comments about the study?

_____

_____

Please list below any university courses you have taken in mathematics, statistics, or logic.

_____

_____

APPENDIX B

Select the response which best describes your feelings on this statement in accordance with the following scale.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

1. Most psychology experiments are worthless since even the most carefully controlled experiments lead to inconclusive results.

2. Through experimentation psychologists have made a real contribution to the understanding of men.

3. Psychologists would be better advised to forget the laboratory, and go into the field where the "real people and problems" are.

4. Many of the questions asked in testing are personal and are none of the experimenter's business.

5. Given a free choice, most students would be willing to volunteer for experiments.

6. Many experimenters are smug and take a pretty high-handed attitude with subjects.

7. Most experiments in psychology are concerned with trivial observations of artificial behavior.

8. Tests and other experimental manipulations are generally not reliable measures of personality and behavior.

9. Most experiments deal with such a small segment of behavior that they are meaningless in the broad picture.

10. People generally express their real feelings on psychological tests.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

11. Psychology experiments are fun but do not prove anything.

12. Human behavior is too complex to cut up and study piece by piece in the laboratory.

13. Most people would say that their experience as a subject in psychological experiments was favourable.

14. When an individual signs up for an experiment, it involves a commitment to do what is asked to the best of his ability.

15. Most students participate willingly in experiments.

16. People rarely express their "real" selves in psychology experiments.

17. Experiments in psychology have no value because of the inherent diversity of man and his environment.

18. Many experimenters ask too much from their subjects.

19. Experiments are nothing but "busy work" for psychologists.

20. Psychology experiments are too time consuming.

21. Some experimenters just seem to be waiting for the subjects to make fools of themselves.

22. As a matter of personal pride, most individuals would try to do their best when acting as a subject.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

23. Experimentation is of no practical value in the understanding of the fundamental causes of behavior.

24. The psychological journals are mostly filled with unimportant trivia.

25. It doesn't matter too much what subjects do; the experimenter usually manipulates the data to prove his hypothesis anyway.

26. Psychological tests are generally reliable measures of personality.

27. Laboratory studies in psychology are too artificial to produce valid data.

28. Most students are "good" subjects, that is, they perform well in their role as experimental subjects.

29. Many subjects in psychological experiments go through the motions without really trying.

30. The experimental method can be used effectively in the study of human behavior.

31. Subjects in most psychology experiments are treated with respect.

32. The experimental approach to psychology has been both fruitful and helpful in understanding human nature.

33. Most experimenters are considerate and polite in their treatment of subjects.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

34. Participation in psychology experiments is not a great imposition on students.

35. Psychologists sometimes forget that subjects are still human beings.

36. Through psychological tests and experiments psychologists have acquired the knowledge to predict behavior in many real life situations.

37. Most students follow the experimenter's instructions carefully so that they will be able to perform as a good subject.

38. Laboratory studies in psychology have contributed significantly to the knowledge of mankind.

39. The complexity of individuals make it necessary to study human behavior under controlled conditions.

40. From experiments, psychologists can validly generalize to the population at large.

41. Subjects in most psychology experiments are treated as guinea pigs.

42. Many students do not cooperate and therefore make poor subjects.

43. Psychology has proven its worth as an experimental science.

44. Any minor discomfort that subjects may go through such as electric shock, embarrassment, etc., is worth it in the long run.

45. Psychological data is useless because its interpretation is based on the manipulation of statistics.

46. Many students feel a responsibility to cooperate in any way possible in the pursuit of knowledge.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

47. Subjects frequently feel manipulated by the experimenter.

48. Participation in psychological experiments is a waste of the students' time.

49. Students should not be asked to give up their time to serve as subjects.

50. College students tend to share with experimenters the hope that the study in which they are participating will in some material way contribute to science.

51. Subjects in psychology experiments are "contributors to science".

52. Experiments in psychology almost always involve deception or "tricking" the subject in some way.

53. I don't usually bother to analyze and explain people's behavior.

54. Once I have figured out a single cause for a person's behavior I don't usually go any further.

55. I believe it is important to analyze and understand our own thinking processes.

56. I think a lot about the influence that I have on other people's behavior.

57. I have found that the relationships between a person's attitudes, beliefs, and character traits are usually simple and straightforward.

58. If I see people behaving in a really strange or unusual manner I usually put it down to the fact that they are strange or unusual people and don't bother to explain it any further.

59. I have thought a lot about the family background and personal history of people who are close to me, in order to understand why they are the sort of people they are.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

60. I don't enjoy getting into discussions where the causes for people's behavior are being talked over.

61. I have found that the causes for people's behavior are usually complex rather than simple.

62. I am very interested in understanding how my own thinking works when I make judgments about people or attach causes to their behavior.

63. I think very little about the different ways that people influence each other.

64. To understand a person's personality/behavior I have found it is important to know how that person's attitudes, belief, and character traits fit together.

65. When I try to explain other people's behavior I concentrate on the person and don't worry too much about all the existing external factors that might be affecting them.

66. I have often found that the basic cause for a person's behavior is located far back in time.

67. I really enjoy analyzing the reasons or causes for people's behavior.

68. I usually find that complicated explanations for people's behavior are confusing rather than helpful.

69. I give little thought to how my thinking works in the process of understanding or explaining people's behavior.

70. I think very little about the influence that other people have on my behavior.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

71. I have thought a lot about the way that different parts of my personality influence other parts (e.g., beliefs affecting attitudes or attitudes affecting character traits).

72. I think a lot about the influence that society has on other people.

73. When I analyze a person's behavior I often find the causes form a chain that goes back in time, sometimes for years.

74. I am not really curious about human behavior.

75. I prefer simple rather than complex explanations for people's behavior.

76. When the reasons I give for my own behavior are different from someone else's, this often makes me think about the thinking processes that lead to my explanations.

77. I believe that to understand a person you need to understand the people who that person has close contact with.

78. I tend to take people's behavior at face value and not worry about the inner causes for their behavior.

79. I think a lot about the influence that society has on my behavior and personality.

80. I have thought very little about my own family background and personal history in order to understand why I am the sort of person I am.

81. I would prefer complex to simple problems.

82. I like to have the responsibility of handling a situation that requires a lot of thinking.

83. Thinking is not my idea of fun.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

84. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

85. I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something.

86. I find satisfaction in deliberating hard and for long hours.

87. I only think as hard as I have to.

88. I prefer to think about small, daily projects to long-term ones.

89. I like tasks that require little thought once I've learned them.

90. The idea of relying on thought to make my way to the top appeals to me.

91. I really enjoy a task that involves coming up with new solutions to problems.

92. Learning new ways to think doesn't excite me very much.

93. I prefer my life to be filled with puzzles that I must solve.

94. The notion of thinking abstractly is appealing to me.

95. I would prefer a task that is intellectual, difficult and important to one that is somewhat important but does not require much thought.

96. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.

97. It's enough for me that something gets the job done; I don't care how or why it works.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

98. I usually end up deliberating about issues even when they do not affect me personally.

99. I enjoy being in front of large audiences.

100. When I meet a stranger, I often think that he is better than I am.

101. It is hard for me to start a conversation with strangers.

102. People seem naturally to turn to me when decisions have to be made.

103. I feel secure in social situations.

104. I like to exert my influence over other people.

105. My behavior is usually an expression of my true inner feelings, attitudes, and beliefs.

106. At parties and social gatherings, I do not attempt to do or say things that others will like.

107. I can make impromptu speeches on topics about which I have almost no information.

108. I guess I put on a show to impress or entertain people.

109. When I am uncertain how to act in social situations, I look to the behavior or others for cues.

110. I would probably make a good actor.

111. In a group of people I am rarely the center of attention.

1	2	3	4	5
STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE

112. In different situations and with different people, I often act like very different persons.

113. I am not particularly good at making other people like me.

114. Even if I am not enjoying myself, I often pretend to be having a good time.

115. I'm not always the person I appear to be.

116. I would not change my opinions (or the way I do things) in order to please someone else or win their favor.

117. I have considered being an entertainer.

118. In order to get along and be liked, I tend to be what people expect me to be rather than anything else.

119. I have never been good at games like charades or improvisational acting.

120. I have trouble changing my behavior to suit different people and different situations.

121. At a party I let others keep the jokes and stories going.

122. I feel a bit awkward in company and do not show up quite as well as I should.

123. I can look anyone in the eye and tell a lie with a straight face (if for a right end).

APPENDIX C

Age _____ Sex _____

This booklet contains a number of problems to be solved. The answers to the problems should be recorded in this booklet in the spaces provided.

Please go through the booklet page by page. **Do not go on to the next page without completing the current page and do not return to previous pages.**

Take care in approaching each problem. Your score will be reflected in how close you come to the correct answer. Work alone and please do not ask any questions until all booklets are completed.

A cab was involved in a hit-and-run accident at night. Two companies, the Green and the Blue, operate in the city. You are given the following data:

(i) Although the two companies are roughly equal in size, 85% of the cab accidents in the city involve Green cabs and 15% involve Blue cabs.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

Question: What is the probability (expressed as a percentage) that the cab involved in the accident was Blue rather than Green?

_____%

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. From the descriptions, one man has been randomly selected and is described below.

> Jack is a 45 year-old man. He is married and has four children. He is generally conservative, careful, ambitious, competitive and argumentative. He is interested in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

The probability that Jack is one of the 30 engineers in the sample of 100 is

_____%.

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

Linda is a 31 years old, single, outspoken, and very bright. As an undergraduate she majored in philosophy and was active in the environmental and civil rights movements. A personnel survey showed that of clerical workers in banks (including tellers), fewer than 1% have personality profiles that sound similar to Linda's. If you stood to win $10 if the statement you choose turns out to be true (whether or not the other statement is also true)...

Is it more probable that Linda is a

1	2	3	4	5	6	7
Bank Teller						Bank Teller and a Feminist

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rich and he expresses himself orally very well. He has much feeling and sympathy for other people and enjoys interacting with others. He has a deep moral sense and likes to give of himself to others.

Is it more probable that Tom is now a graduate student in

1	2	3	4	5	6	7
Education						Library Science

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

Bret was a student in social psychology last year. Brief essays were written on the topic of whether there should be more or less funding for research into AIDS. Bret was assigned to write a brief essay arguing for less funding for research into AIDS, an argument his professor had endorsed in class. This is his essay.

> As usual when a new threatening disease is discovered it attracts more media attention than it deserves. Acquired Immune Deficiency Syndrome (AIDS) is comparatively rare and did not warrant the coverage in the media it has received. An enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases. If that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.

Bret is _____ increasing funding for research into AIDS.

1	2	3	4	5	6	7
Strongly Against						Strongly For

On the following scale, please circle the number that indicates your confidence in the judgment you have made:

1	2	3	4	5	6	7
Very Unsure						Very Sure

Perceptions of the Experimental Task

The first part of the experiment is now over. However, the following questions are an important part of this study. They are designed to help me understand your responses to the first part of the study. One of my concerns is to attempt to understand how you perceived each of the problems I presented to you. There are no right or wrong answers to the questions that follow. I am interested in your views of the problems <u>during</u> the experiment.

I don't want you to analyze the problems again. I want you to think back to when you first read the problems. What did you think of the problems, what did you attend to, and what did you base your answers upon? On the next few pages, please tell me honestly what thoughts you had about the experiment.

What do you think was the purpose of this experiment?

_____

_____

_____

Why did you think I asked you to solve these problems?

_____

_____

_____

What quality or skill of yours do you think I was trying to assess with these problems?

_____

_____

_____

Did you feel that the problems really didn't have a correct response? If so, which problems and why?

_____

_____

_____

Did you think there was any trick to answering the problems?

_____

_____

_____

Do you think there was a single correct response or more than one correct response to the cab problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the cab problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

A cab was involved in a hit-and-run accident at night. Two companies, the Green and the Blue, operate in the city. You are given the following data:

(i) Although the two cab companies are roughly equal in size, 85% of the cab accidents in the city involve Green cabs and 15% involve Blue cabs.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

Question: What is the probability (expressed as a percentage) that the cab involved in the accident was Blue rather than Green?

_____%.

Please rate each of the following statements as to their relevance to your arriving at a solution to the cab problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ A cab was involved in a hit-and-run accident at night.

_____ Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs.

_____ A witness identified the cab as a Blue cab.

_____ The witness made correct identifications in 80% of the cases and erred in 20% of the cases.

Please read the following two interpretations of the cab problem (reproduced below):

1. Since the witness was 80% accurate in the court test, the probability that the cab which committed this crime is Blue is about 80%.

2. Since 85% of the cab accidents in the city involve Green cabs, the probability that the cab which committed this crime is Blue is about 15%.

Would you say that you:

1	2	3	4	5	6	7
Strongly Endorse Interpretation 1						Strongly Endorse Interpretation 2

A cab was involved in a hit-and-run accident at night. Two companies, the Green and the Blue, operate in the city. You are given the following data:

(i) Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

What is the probability (expressed as a percentage) that the cab involved in the accident was Blue rather than Green?

____%

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a high number for the Cab problem (reproduced below).

_____

_____

_____

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a low number for the Cab problem (reproduced below).

_____

_____

_____

A cab was involved in a hit-and-run accident at night. Two companies, the Green and the Blue, operate in the city. You are given the following data:

(i) Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs.

(ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

What is the probability (expressed as a percentage) that the cab involved in the accident was Blue rather than Green?

____%

For the Cab problem, I am interested in your recollection of the details of the problem. In the space below, please attempt to write down the problem in your own words. Do the best you can but do not look back at the problem.

_____

_____

_____

_____

_____

Do you think there was a single correct response or more than one correct response to the engineer/lawyer problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the engineer/lawyer problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. From the descriptions, one man has been randomly selected and is described below.

> **Jack is a 45 year-old man. He is married and has four children. He is generally conservative, careful, ambitious, competitive and argumentative. He is interested in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.**

The probability that Jack is one of the 30 engineers in the sample of 100 is

_____%.

Please rate each of the following statements as to their relevance to your arriving at a solution to the engineer/lawyer problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant						Extremely relevant

_____ A panel of psychologists have interviewed and administered personality tests and prepared thumbnail sketches.

_____ 30 engineers and 70 lawyers, all successful

_____ Jack is a 45 year-old man, married with four children.

_____ He is generally conservative, careful, ambitious, competitive, and argumentative.

_____ He is interested in political and social issues.

_____ He spends most of his free time on his many hobbies: home carpentry, sailing, and mathematical puzzles.

Please read the following two interpretations of the Jack problem (reproduced below):

1. In order to determine whether Jack is an engineer or lawyer, it is important to look for clues in the description of his personality. The description of Jack seems consistent with that of an engineer, so I judged him to be more likely to be an engineer.

2. In order to determine whether Jack is an engineer or lawyer, it is important to consider the sample that Jack was drawn from. Since most of the individuals in the sample were lawyers, Jack is probably a lawyer.

Would you say that you:

1	2	3	4	5	6	7
Strongly Endorse Interpretation 1						Strongly Endorse Interpretation 2

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. From the descriptions, one man has been randomly selected and is described below.

Jack is a 45 year-old man. He is married and has four children. He is generally conservative, careful, ambitious, competitive and argumentative. He is interested in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

The probability that Jack is one of the 30 engineers in the sample of 100 is

_____%.

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a high number for the engineer/lawyer problem (reproduced below).

_____

_____

_____

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a low number for the engineer/lawyer problem (reproduced below).

_____

_____

_____

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. From the descriptions, one man has been randomly selected and is described below.

Jack is a 45 year-old man. He is married and has four children. He is generally conservative, careful, ambitious, competitive and argumentative. He is interested in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

The probability that Jack is one of the 30 engineers in the sample of 100 is

____%.

For the engineer/lawyer problem, I am interested in your recollection of the details of the problem. In the space below, please attempt to write down the problem in your own words. Do the best you can but do not look back at the problem.

_____

_____

_____

_____

_____

Do you think there was a single correct response or more than one correct response to the Linda problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the Linda problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

Linda is a 31 years old, single, outspoken, and very bright. As an undergraduate she majored in philosophy and was active in the enviornmental and civil rights movements. A personnel survey showed that of clerical workers in banks (including tellers), fewer than 1% have personality profiles that sound similar to Linda's. If you stood to win $10 if the statement you choose turns out to be true (whether or not the other statement is also true)...

Is it more probable that Linda is a

1 2 3 4 5 6 7

Bank Teller — **Bank Teller and a Feminist**

Please rate each of the following statements as to their relevance to your arriving at a solution to the Linda problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ Linda is a 31 year-old, single, outspoken, and very bright.

_____ As an undergraduate she majored in philosophy and was active in the environmental and civil rights movements.

_____ A personnel survey showed that of clerical workers in banks (including tellers), fewer than 1% have personality profiles that sound similar to Linda's.

Please read the following two interpretations of the Linda problem (reproduced below):

1. Linda is a bank teller because the probability that Linda is both a bank teller and an active feminist must be smaller than the probability she is a bank teller.

2. Linda is a bank teller and an active feminist because Linda resembles a bank teller who is active in the feminist movement more than she resembles a bank teller.

Would you say that you:

1	2	3	4	5	6	7
Strongly Endorse Interpretation 1						Strongly Endorse Interpretation 2

Linda is a 31 years old, single, outspoken, and very bright. As an undergraduate she majored in philosophy and was active in the environmental and civil rights movements. A personnel survey showed that of clerical workers in banks (including tellers), fewer than 1% have personality profiles that sound similar to Linda's. If you stood to win $10 if the statement you choose turns out to be true (whether or not the other statement is also true)...

Is it more probable that Linda is a

1	2	3	4	5	6	7
Bank Teller						Bank Teller and a Feminist

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a high number for the Linda problem (reproduced below).

_____

_____

_____

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a low number for the Linda problem (reproduced below).

_____

_____

_____

Linda is a 31 years old, single, outspoken, and very bright. As an undergraduate she majored in philosophy and was active in the environmental and civil rights movements. A personnel survey showed that of clerical workers in banks (including tellers) fewer than 1% have personality profiles that sound similar to Linda's. If you stood to win $10 if the statement you choose turns out to be true (whether or not the other statement is also true)...

Is it more probable that Linda is a

1	2	3	4	5	6	7
Bank Teller						Bank Teller and a Feminist

For the Linda problem, I am interested in your recollection of the details of the problem. In the space below, please attempt to write down the problem in your own words. Do the best you can but do not look back at the problem.

_____

_____

_____

_____

_____

Do you think there was a single correct response or more than one correct response to the Tom W. problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the Tom W. problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

> **Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rich and he expresses himself orally very well. He has much feeling and sympathy for other people and enjoys interacting with others. He has a deep moral sense and likes to give of himself to others.**

Is it more probable that Tom is now a graduate student in

1 2 3 4 5 6 7

Education — Library Science

Please rate each of the following statements as to their relevance to your arriving at a solution to the Tom W. problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant						Extremely relevant

_____ The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

_____ Tom W. is of high intelligence, although lacking in true creativity.

_____ He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place.

_____ His writing is rich and he expresses himself orally very well.

_____ He has much feeling and sympathy for other people and enjoys interacting with others.

_____ He has a deep moral sense and likes to give of himself to others.

Please read the following two interpretations of the Tom W. problem (reproduced below):

1. In order to determine whether Tom W. is a graduate student in education or library science, it is important to look for clues in the description of his personality. The description of Tom W. seems consistent with that of a library science graduate student, so I judged him to be more likely to be in library science.

2. In order to determine whether Tom W. is a graduate student in education or library science, it is important to consider the number of graduate students enrolled in each program. Since there are more graduate students in education, Tom W. is most probably a graduate student in education.

Would you say that you:

1	2	3	4	5	6	7
Strongly Endorse Interpretation 1						Strongly Endorse Interpretation 2

The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

> **Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rich and he expresses himself orally very well. He has much feeling and sympathy for other people and enjoys interacting with others. He has a deep moral sense and likes to give of himself to others.**

Is it more probable that Tom is now a graduate student in

1	2	3	4	5	6	7
Education						Library Science

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a high number for the Tom W. problem (reproduced below).

_____

_____

_____

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a low number for the Tom W. problem (reproduced below).

_____

_____

_____

The following is the description of Tom W. written by a psychologist who based his description on projective tests administered when Tom W. was in high school.

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rich and he expresses himself orally very well. He has much feeling and sympathy for other people and enjoys interacting with others. He has a deep moral sense and likes to give of himself to others.

Is it more probable that Tom is now a graduate student in

1	2	3	4	5	6	7
Education						Library Science

For the Tom W. problem, I am interested in your recollection of the details of the problem. In the space below, please attempt to write down the problem in your own words. Do the best you can but do not look back at the problem.

_____

_____

_____

_____

_____

Do you think there was a single correct response or more than one correct response to the Bret problem (reproduced below)? Please check the appropriate box.

|__| |__| |__| |__| |__| |__| |__|

A single correct response — Many correct responses

Do you think that the Bret problem demanded an intuitive and personal judgment based on your life experience or a formal analysis based on the application of mathematics and rules of logic?

|__| |__| |__| |__| |__| |__| |__|

Intuitive judgment — Formal analysis

Bret was a student in social psychology last year. Brief essays were written on the topic of whether there should be more or less funding for research into AIDS. Bret was assigned to write a brief essay arguing for less funding for research into AIDS, an argument his professor had endorsed in class. This is his essay.

> **As usual when a new threatening disease is discovered it attracts more media attention than it deserves. Acquired Immune Deficiency Syndrome (AIDS) is comparatively rare and did not warrant the coverage in the media it has received. An enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases. If that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.**

Bret is _____ increasing funding for research into AIDS.

1 2 3 4 5 6 7

Strongly Against — Strongly For

Please rate each of the following statements as to their relevance to your arriving at a solution to the Bret problem according to the following scale.

1	2	3	4	5	6	7
Not at all relevant					Extremely relevant	

_____ Bret was assigned to write a brief essay arguing for less funding for research into AIDS, an argument his professor had endorsed in class.

_____ Bret wrote that AIDS is comparatively rare and did not warrant the coverage in the media it has received.

_____ Bret wrote that an enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases.

_____ Bret wrote that if that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.

Please read the following two interpretations of the Bret problem (reproduced below):

1. In order to determine Bret's true attitude toward funding AIDS research, it is important to look for clues in the essay he wrote. No one could write an essay like that unless they believed that funding for AIDS research should be less.

2. In order to determine Bret's true attitude toward funding for AIDS research, one must remember that Bret was assigned by the professor to write an essay arguing against funding for AIDS research. The essay tells us nothing about what he really thinks.

Would you say that you:

1	2	3	4	5	6	7
Strongly Endorse Interpretation 1						Strongly Endorse Interpretation 2

Bret was a student in social psychology last year. Brief essays were written on the topic of whether there should be more or less funding for research into AIDS. Bret was assigned to write a brief essay arguing for less funding for research into AIDS, an argument his professor had endorsed in class. This is his essay.

> **As usual when a new threatening disease is discovered it attracts more media attention than it deserves. Acquired Immune Deficiency Syndrome (AIDS) is comparatively rare and did not warrant the coverage in the media it has received. An enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases. If that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.**

Bret is _____ increasing funding for research into AIDS.

1	2	3	4	5	6	7
Strongly Against						Strongly For

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a high number for the Bret problem (reproduced below).

_____

_____

_____

Please write any relevant reasons, considerations, arguments, etc., that would justify recording a low number for the Bret problem (reproduced below).

_____

_____

_____

Bret was a student in social psychology last year. Brief essays were written on the topic of whether there should be more or less funding for research into AIDS. Bret was assigned to write a brief essay arguing for less funding for research into AIDS, an argument his professor had endorsed in class. This is his essay.

> As usual when a new threatening disease is discovered it attracts more media attention than it deserves. Acquired Immune Deficiency Syndrome (AIDS) is comparatively rare and did not warrant the coverage in the media it has received. An enormous amount of energy and costly research is directed towards the investigation of AIDS at the expense of more important life-threatening diseases. If that same energy and investment were directed toward the control of more common conditions patients would achieve greater benefit.

Bret is _____ increasing funding for research into AIDS.

1	2	3	4	5	6	7
Strongly Against						Strongly For

For the Bret problem, I am interested in your recollection of the details of the problem. In the space below, please attempt to write down the problem in your own words. Do the best you can but do not look back at the problem.

Have you seen any of the problems before?

_____

If so, which problems and where did you see them before?

_____

_____

Do you have any other comments about the study?

_____

_____

Please list below any university courses you have taken in mathematics, statistics, or logic.

_____

_____