INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

NOTE TO USERS

This reproduction is the best copy available.

UMI®

Université d'Ottawa • University of Ottawa

INFORMATION REPRESENTATION, PROBLEM FORMAT, AND MENTAL ALGORITHMS IN PROBABILISTIC REASONING

By

Owen Helmkay
School of Psychology
University of Ottawa

Doctoral dissertation submitted to the
School of Graduate Studies of the University of Ottawa

In partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Clinical Psychology

August, 1999

© Owen Helmkay, Ottawa, Canada 1999

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services
Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

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

0-612-66153-9

Canada

To Aileen, Annika, and Evan.

Your love and support during the final stages of writing this thesis has allowed me to look forward to a future renewed with love and meaning.

Acknowledgements

I would like to express my sincerest gratitude to my thesis advisor, Dr. Pierre Mercier, for all the contributions he has made throughout the development, implementation, and writing of this thesis and for all the support he has shown me over the course of my doctoral studies. As a supervisor and mentor, I can imagine no better person.

I would also like to thank the members of my dissertation committee: Drs. Alain Desrochers, Luc Pelletier, and Catherine Plowright for their helpful feedback and suggestions in conducting the final phases of this research. I would also like to thank Dr. Henry Markovits, who served as the external examiner, for his comments and challenging questions during the thesis defense. Your contributions are greatly appreciated.

I would also like to take this opportunity to thank my family and friends for their encouragement and support while I was busy with my studies. Thank you all for your understanding!

INFORMATION REPRESENTATION, PROBLEM FORMAT, AND MENTAL ALGORITHMS IN PROBABILISTIC REASONING

Owen Helmkay

Abstract

The purpose of this thesis was to better understand when people use and neglect base rate information. In four experiments, university students made probabilistic judgements for inferential reasoning tasks that were modelled after Tversky and Kahneman's taxi-cab problem (Tversky & Kahneman, 1982). In all, four theoretical perspectives were used to test predicted outcomes with observed responses and normative expectations. Experiment 1 found no differences between the trial-by-trial (on-line) format (Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993) and the traditional word problem (off-line) format of Tversky and Kahneman. Here the pattern of judgements was nearly identical across conditions and yet different from normative expectations. Both formats showed a strong sensitivity to base rates and the individuating information. This was contrary to the predictions based on the heuristics and biases program (Kahneman, Slovic, & Tversky, 1982) and also contrary to the predictions made by the frequentist approach (Gigerenzer, 1998; Gigerenzer & Hoffrage, 1995). Experiments 2 and 3 were designed to test the maxims of conversational conventions (Grice, 1975; Hilton, 1995). Experiment 2 was run entirely on-line and found no differences in mean probability judgements between the different conditional (individuating) sources of information. Participants relied equally heavily on a human or non-human source and also were sensitive to the changes made in base rate information. Experiment 3 was run entirely off-line and specifically tested the attribute of ambiguity. When the individuating source of information was very ambiguous, then judgements were based almost solely on the base rate information. None of the theories tested to this point were able to accurately predict the patterns of judgements made in all these conditions. What the results of these three experiments did suggest was that how the information is presented (e.g., probabilities vs. frequencies) and the context in which it is presented (i.e., how clearly the information is presented) are both important. These two factors combined suggest that a more general cognitive mechanism, one that is a function of the complexity of the judgement task and the amount of cognitive work that is required to make the judgement could best account for the data. This was tested and supported in Experiment 4. Normative responses can be elicited in judgement tasks that present only the essential probabilistic information in a format that is clear, unambiguous, and that requires few mental operations.

Table of Contents

List of Figures

List of Tables

INFORMATION REPRESENTATION, PROBLEM FORMAT, AND MENTAL ALGORITHMS IN PROBABILISTIC REASONING

Introduction

The area of investigation commonly referred to as Judgement (or Decision Making) Under Uncertainty is interdisciplinary in its scope, drawing on research from the areas of psychology, mathematics, philosophy, economics, medicine, and forecasting among others. Its history is a long and well established one, with scientific literature on the topic dating back to the 15th and 16th centuries (see reviews by Dellarosa (1988) and Gigerenzer and Murray (1987)). More recently, the topic of rational thought has been implicitly defined by the laws of probability theory with one major thrust coming from the heuristics-and-biases program developed by Daniel Kahneman and Amos Tversky (Kahneman & Tversky, 1972; 1973; Tversky & Kahneman, 1971; 1973; 1974a; 1980; 1982c; 1982e; 1982f). This program laid the foundation for a theorised heuristic process of probabilistic reasoning (e.g., representativeness, availability, simulation, and anchoring and adjustment) and was the impetus for many studies investigating such phenomena as the conjunction fallacy (Gigerenzer, 1991b; Hertwig & Chase, 1995; Hertwig & Gigerenzer, 1995; Tversky & Kahneman, 1983), overconfidence (Gigerenzer, 1991b; Lichtenstein & Fischhoff, 1977), and perhaps the most well known and researched, the base-rate fallacy (Ajzen, 1977; Bar-Hillel, 1980; 1983; 1990; Carroll & Siegler, 1977; Christensen-Szalanski & Beach, 1982b; Christensen-Szalanski & Bushyhead, 1981; Cosmides & Tooby, 1996; Gigerenzer, 1991b; Gigerenzer & Hoffrage, 1995; Tversky & Kahneman, 1982c; 1982e). One of the classic tasks illustrating the base-rate fallacy is the cab problem described by Tversky and Kahneman (1982c, p.156):

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and Blue, operate in the city. You are given the following data: (a) 85% of the cabs in the city are Green and 15% are Blue. (b) A witness identified the cab as

> Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each of the two colours 80% of the time and failed 20% of the time.
>
> What is the probability that the cab involved in the accident was Blue rather than Green?

In this problem, the correct answer is 0.41. When the problem is stated as above, this answer can be arrived at by solving Bayes' theorem which takes into account two types of information: 1- the witness reliability and 2- the proportion of blue cabs in the city (see Equations 1 and 2 in the section on Normative Models of Probabilistic Reasoning below for a full explanation). However, the vast majority of people tested with this problem provide an answer of 0.80 or so, suggesting that they ignore the proportion of blue cabs. This is the phenomenon referred to as the base-rate fallacy. Research with this and many other similar problems resulted in plenty of evidence showing that participants made judgements that often completely ignored base rates and always failed to use base rates to their full extent (Bar-Hillel, 1980; 1983; 1984; 1990; Kahneman & Tversky, 1982b; Tversky & Kahneman, 1980). Instead, people appear to rely on the diagnostic or individuating information at hand: The witness reliability.

In 1980, the bulk of the studies involving the manipulation of base rates lead Bar-Hillel to the conclusion that "the genuineness, the robustness, and the generality of the base-rate fallacy are matters of established fact" (p. 215). In that same paper, she also concluded that the use of base rates was a function of their specificity and relevance. For instance, in the cab problem, two pieces of information are presented: the base rate and the witness' testimony. She argued that the base rate was not seen as relevant to the judgement. Thus, Bar-Hillel (1980) presented subjects with other problems where both the base rate and the indicant information were equally relevant. For example, the eyewitness testimony was replaced by a pedestrian who heard an intercom coming through the cab window. A police investigation indicated that intercoms were installed in 80% of the Green cabs and in 20% of the Blue cabs. The pattern of judgements to this problem showed that the indicant information did not dominate the base rates - that the distribution of responses were relatively flat (compared to the original cab problem) with the median estimate at 48% which is quite close to the normative judgement of 41%. In spite of these results, we are still left today without a clear understanding of all the conditions under which people pay attention to base rates and how they use that information when it is presented to them (see the review of the base rate literature presented by Koehler (1996)).

Recently, Gigerenzer (Gigerenzer, 1991b; Gigerenzer & Hoffrage, 1995; Gigerenzer & Murray, 1987) proposed that part of the reason for the non-normative (Bayesian) use of base rates in probabilistic reasoning is due to problem format, that is, how the information is presented. In most probabilistic reasoning tasks such as the cab problem, the information has been presented in what is termed a single-event probability format. That is, the problem statement refers to one particular event (e.g. a cab involved in an accident) and the numerical information available is expressed in proportions or percentages. Based on work from evolutionary theory (e.g., Cosmides & Tooby, 1996) and on the automatic processing of frequency information (e.g., Hasher & Zacks, 1979), Gigerenzer proposes that presenting information in this manner is "unnatural" in that the cognitive system is not structured to process this type of information on the fly, without the help of paper and pencil or a calculator. This makes the problem harder to solve and, most importantly, prevents the researcher from observing what the brain is fully capable of. Gigerenzer argued that presenting similar problems in a frequentist format, whereby the information is expressed as relative frequencies (e.g. 15 cabs out of 100 are Blue) would more easily elicit the natural information processing mechanism that the human brain has evolved, thereby resulting in more normative answers. When Gigerenzer and his colleagues presented the information in a frequentist format (Gigerenzer & Hoffrage, 1995), they did observe more normative answers. Unfortunately, Gigerenzer's methodology gives rise to a new difficulty in interpreting the results globally. The question answered by Gigerenzer's research participants was not worded the same as that used in the classic problem formulation. The most normative answers were obtained when the question was posed as: "How many cabs that the witness reports to be blue are actually Blue? ____ out of ____. " This leaves open the possibility that the different question wording *induces* a different problem-solving strategy. If the cognitive processes triggered by the wording of the single-event problem format are different from those triggered by the wording of the relative frequency presentation, quite apart from the format of the numbers themselves, then the two research programs become difficult to compare. It could be that people solve probability problems using heuristics or computing relative frequencies or both, but neither research program tells us unambiguously which one is used when. To resolve the issue, experiments are needed that frame the problem question in a manner identical to that used in the research inspired by Tversky, Kahneman and Bar-Hillel and yet allow people to solve the problem any which way they can, including via the computation of relative frequencies.

An alternative method of presenting probabilistic problems in a frequency format can be done graphically, on-line, with the information presented trial-by-trial. A trial-by-trial version of the cab problem would involve presenting a series of animated cabs crossing a computer screen and becoming involved in accidents with a bystander witnessing the scene and providing testimony as to the colour of the cab. Imagine for example that, in a presentation of 40 cabs, 6 are Blue and 34 are Green. Of the 6 Blue cabs, the witness correctly reports blue in 5 cases (when the cabs are actually Blue) and incorrectly reports green in 1 case. Of the 34 Green cabs, the witness correctly reports green in 27 cases (when the cabs are actually Green) and incorrectly reports blue in 7 cases. After these 40 trials, a forty-first animation presents an unidentified cab, black with a big yellow question mark adorning it, which ends up in an accident. The witness makes a statement as to the actual, hidden, colour of the cab, and the research participant is asked a probabilistic question similar in structure but even more explicit than that used by Bar-Hillel (1980; 1984): "What is the probability that the cab involved in the accident is Blue, given that the witness said it was blue?" Empirical evidence from research on contingency judgements (e.g., Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993) suggests that people can make accurate probabilistic judgements in such on-line tasks (Cheng, 1995; Shanks, 1985). In the Baker et al. experiments for instance, animated tanks were presented on a computer screen, passing through a minefield with or without camouflage. Trial-by-trial, participants would observe the tank to either explode or pass safely through the minefield. After 40 trials, participants were asked to assess the relative safety of the tank when it was camouflaged as compared to when it was not. These contingency judgements were made on a scale ranging from -100 to +100, where -100 meant that the camouflage was extremely dangerous, +100 meant that the camouflage was extremely safe, and 0 meant that the camouflage was neither safer nor more dangerous than no camouflage. The normative answer to this contingency judgement is the ΔP coefficient (Allan, 1980) which is the difference between the conditional probability of safety given camouflage and the conditional probability of safety in the absence of camouflage. Neither Baker et al. nor Cheng nor Shanks, argue that people arrive at contingency judgements by computing ΔP but all obtained a fair degree of correspondence between participants' estimated contingency judgements and normative expectations. Thus the results of these experiments suggest that the cognitive mechanism underlying them can often produce judgements that are close to normative expectations. If research participants were able to accurately discriminate contingencies involving two conditional probabilities in these experiments where the information was presented on-line, then importing the on-line attribute into the cab problem while preserving the exact wording of the question might accomplish the double objective of producing more normative answers and clarifying what makes this possible. This is the goal of the first experiment.

Overview of the Experiments

Based on (a) Gigerenzer's theoretical argument that frequentist information is a more natural representation of probabilistic information, and (b) evidence from trial-by-trial studies of contingency judgement showing that normative probability judgement is at least possible, Experiment 1, was developed as an on-line / off-line format of the cab problem. The on-line format maintained the integrity of the problem while being more ecologically sensitive to how probability information occurs outside of the laboratory. The ecological validity of this task was deemed better than that of either the classical problem or of other frequentist versions used so far. The off-line format was identical to the methodology used by Kahneman and Tversky.

Contrary to expectation, the results of Experiment 1 did not match those obtained by Gigerenzer and colleagues in spite of the ecologically valid frequentist format used to present the information. Participants produced similarly biased judgements in the on-line as in the off-line condition. Since Experiment 1 and Gigerenzer's work clearly differ in how the problem question was formulated, a detailed examination of the question wording was undertaken. A theoretical context for such an analysis is provided by the conversational inference framework (Hilton, 1995b; Schwarz, Strack, Hilton, & Naderer, 1991). This approach suggests that judgmental inferences made in many areas of experimental psychology must be considered in the context in which the information is presented. According to this approach, biased judgements in the cab problem may not stem from a frequentist versus standard probability formulation, but may be a function of conversational maxims attributable to the witness' testimony. Using Grice's (1975) conversational conventions analysis, Hilton (1995b) proposes that many of the biases found in psychological research may be less attributable to cognitive factors such as representativeness and base-rate neglect (Kahneman & Tversky, 1973) than to inferences about the social context of the problem which are guided by conversational assumptions. Of the four conversational maxims postulated by Grice (1975), the most important ones as they pertain to the cab problem involve the maxims of quality, relation, and manner. The maxim of quality pertains to the truth-value of the information with the corresponding attributes of sincerity, honesty, reliability, and competence of the sources (the witness' testimony and the base rate). The maxim of relation prescribes that the information presented is relevant to the goals of the judgement task. The maxim of manner pertains to the brevity, clarity, orderliness, and freedom from ambiguity of the information presented. The witness reliability is quantified explicitly in the cab problem. Thus participants would be left to assume that the witness is simultaneously very sincere, honest and competent. This information competes with the base rate information that is quantified in the problem as the number of Green and Blue cabs in the city. The base rate, however, may be seen as less relevant than the witness reliability information. In addition, sincerity, honesty and competence are not relevant attributes for the base rate information. Reliability is the only attribute that might matter for the base rate and yet no statement is made about it. Thus, applying these maxims to the cab problem, and assuming a desire to co-operate from the participant, a conversational analysis predicts an over-reliance on the witness information and under-utilisation of base rates (Hilton, 1995b).

Based on this theoretical framework, a new conditional situation was devised that would modify the balance of these conversational attributes. Unlike Experiment 1, the new condition devised for Experiment 2 presented cabs that were involved in accidents on some but not all of the trials. This manipulation relies on the accidental status for the cab and the corresponding question asks, "What is the probability that the cab was Blue given that it was involved in an accident?" In this new condition, no witness was present at all. Participants saw Green and Blue cabs involved in hit-and-run accidents on some occasions and cabs passing safely by on other occasions. The proportion of Blue cabs involved in accidents directly corresponded to the proportion of Blue cabs that the witness reported to be blue in the standard task. This accidental status condition and the witness condition match one another in their conceptual structure. The only difference lies in the source of the conditional statement. According to the conversational inference hypothesis, the witness as the source of the conditional is expected to produce more biased judgements whereas the accidental status as the conditional source is expected to elicit judgements closer to normative expectations. This is because the accidental status may remove the influence of three of the four attributes from the maxim of quality (sincerity, honesty and competence), leaving the attribute of reliability with both the accidental status and witness conditions. Now, the accidental status and the base rate may be perceived as contributing more equally as sources of probabilistic information. The reliability of the two sources is never mentioned explicitly yet it could be computed equally well from the trial-by-trial data for both the accidental status and the base rate if the participants wished to do so. This was tested on-line in Experiment 2 and off-line in Experiment 3. Neither experiment produced more normative answers as a main effect of the accidental status condition.

In Experiment 3, the maxim of manner, specifically the property of ambiguity, was tested by creating ambiguous and unambiguous formulations of the problem. The information in the cab problem reports that the witness is correct 80% of the time, leaving it to the problem-solver to construct a representation of this information. The correct formulation corresponds to 80% of the Blue cabs being identified as blue. It would be easy, however, for the problem-solver to misconstrue this as meaning that 80% of stated blue cabs are really Blue. Those two proportions differ radically. To test this, two problem formulations, one ambiguous (base rate left implicit) and one unambiguous (base rate explicitly stated), were developed (see details in Experiment 3). This experiment showed that judgements were more normative in the condition where the problem used the accidental status of the cab as one source of information and the formulation was ambiguous. Neither a frequentist nor a simple conversational analysis approach was sufficient to explain this interaction. Rather, the pattern of individual responses suggested that, free from a strong influence of the maxim of quality but confronted with a high level of problem difficulty combined with ambiguous information, many participants tried different solutions to the problem.

Pursuing this, Experiment 4 manipulated problem difficulty off-line by presenting problems that could be solved in either just one, two, or three operational steps, and showed that this dimension of problem-solving to be the best single predictor of how normative the judgements can be. But before delving into the details of these four experiments, let us first examine the relevant theory and research more fully.

Normative Model of Probabilistic Reasoning

Bayes' Theorem is a dynamic theory of normative opinion revision that has been used for updating prior probabilities as more information is obtained (Wright, 1984). In its simplest form, Bayes' Theorem states that the posterior opinion that a hypothesis is true is the product of the prior opinion multiplied by the likelihood of the obtained data given that the hypothesis is true. Therefore, for binary hypotheses, Bayes' Theorem can be expressed mathematically as:

$$P(H \mid E) = \frac{P(E \mid H) \bullet P(H)}{P(E \mid H) \bullet P(H) + P(E \mid \overline{H}) \bullet P(\overline{H})} \quad (1)$$

where H and E represent the Hypothesis and Evidence respectively. $P(H)$ and $P(\overline{H})$ are the prior probabilities and refer to the probabilities of the truth and falsity of a hypothesis H prior to the collection of additional data. $P(E|H)$ and $P(E|\overline{H})$ are the posterior probabilities and refer to the information value of the evidence if the hypothesis is true and false respectively.

Correctly solving the cab problem in its classic form requires the application of Bayes' Theorem. The first step requires the problem-solver to formulate the information in the problem to match the structure of the model (the Bayesian formulae) and then compute the terms of the equation. Therefore, the answer to the question: "What is the probability that the cab was Blue given the evidence provided by the witness?" can be found by computing:

$$P(B \mid W) = \frac{P(W \mid B) \bullet P(B)}{P(W \mid B) \bullet P(B) + P(W \mid G) \bullet P(G)} \quad (2)$$

where P(B) and P(G) refer to the base rates of Blue and Green cabs in the city: For instance, 15 and 85 respectively. P(W|B) refers to the evidence provided by the witness, namely the probability that the witness correctly identifies the cab as blue when it is Blue (80%), while P(W|G) refers to the probability that the witness incorrectly identifies the cab as blue when it is Green (20%).

Therefore, by Bayes' Theorem, the following judgement can be determined:

$$P(B\mid W) = \frac{(.80)(.15)}{(.80)(.15) + (.20)(.85)} = 0.41 \tag{3}$$

Consequently, the hit-and-run cab is more likely to be Green than Blue despite the accuracy of the witness' report simply because the base rate is greater for the Green cab (.85) than it is for the Blue cab (.15). However, this differs dramatically from the empirical results found by Tversky and Kahneman (1982c). In their many studies investigating how individuals arrive at judgements involving the cab problem, they found that participants consistently made median and modal judgements of .80, judgements that corresponded to the accuracy of the witness' testimony. These and other studies were the impetus for Tversky and Kahneman to propose a heuristic style of reasoning under conditions of uncertainty.

Cognitive Heuristics

A heuristic is a rule of thumb or shortcut that is meant to aid the problem-solver to arrive at a judgement or make a decision. It reduces the complex problem-solving process to operations that are simpler than the formal solution (Tversky & Kahneman, 1974). For example, if someone could not determine the correct answer on a multiple choice question, then a heuristic might be that the test-taker use the following rule: if in doubt, choose answer C. Although they are simple, heuristics suffer the obvious shortcoming of not always producing the correct answer - biases are inherent in their use. A central claim of Tversky and Kahneman's research program (1974) is that decisions made under uncertainty most often rely on heuristics. A brief illustration of some of the most notable cognitive heuristics, namely representativeness, availability, and anchoring and adjustment, are presented here.

Representativeness

Representativeness involves the "application of resemblance or 'goodness of fit' criteria to problems of categorisation" (Nisbett & Ross, 1980, p. 24). Representativeness may be used when making judgements involving probability comparisons (e.g., what is the probability that A is associated with B, or that A originated from B). Although representativeness may be a useful rule of thumb when making predictions about the association between two events or items, the decision-maker seems to be unaware or insensitive to a number of factors that affect the actual probability of occurrence. The following problem has elicited a pattern of judgements that are believed to be made by representativeness (Kahneman & Tversky, 1973).

Lawyer-Engineer Problem

> 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. Below you will find a single description chosen at random from the 100 available descriptions.
>
> 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 an engineer is: ____________________%

Here, representativeness refers to how closely the description of Jack corresponds to that of the target category, in this case an engineer. The closer the correspondence, the higher the judged response would be. Kahneman and Tversky (1973) found that judgements made with this personality description (referred to as individuating information) were almost the same regardless of the manipulation of the base rates (i.e., the number of engineers and number of lawyers). The results of this research showed a significant but small judgement difference between the low and high base rate conditions. Judgements averaged 50% for the 30/70 engineer/lawyer condition and 55% for the 70/30 engineer/lawyer condition. Only when the personality description information was not included in problem did participants provide judgements that closely matched the base rate information. This lead Kahneman and Tversky to conclude that:

> explicit manipulation of the prior distribution had a minimal effect on subjective probability.... subjects applied their knowledge of the prior only when they were given no specific evidence. As entailed by the representativeness hypothesis, prior probabilities were largely ignored when individuating information was made available (Kahneman & Tversky, 1982a, p. 56).

These experiments were the beginning of a whole line of research that has been referred to as the heuristics-and-biases program (Gigerenzer, 1991b).

Using representativeness to arrive at probabilistic judgements fails to consider a number of normative rules of decision making thus biasing judgements away from normative expectations (Kahneman, Slovic, & Tversky, 1982). One such consideration involves the prior probability of outcomes. This is the tendency to ignore the relative frequency of occurrence of the events (base rates) and judge the situation using primarily the similarity between the instance proposed in the problem and a general reference. Sample size is another consideration. People appear to have a tendency to base judgements on smaller samples with larger odds rather than on larger samples with smaller odds. Another consideration involves misconceptions about chance and what chance occurrences should look like. For example, on six flips of a fair coin, we perceive the sequence of Heads and Tails "H T H T H H" as more like chance than "H H H T T T" (Tversky & Kahneman, 1974).

Availability

The availability heuristic refers to the ease with which an instance of a class of events can be retrieved (Tversky & Kahneman, 1974). It is presumed to be used when evaluating the frequency or likelihood of an event. Consider the following problem used by Tversky and Kahneman (1982a, p. 167):

> The frequency of appearance of letters in the English language was studied. A typical text was selected, and the relative frequency with which various letters of the alphabet appeared in the first and third positions in words was recorded. Words of less than three letters were excluded from the count.

You will be given several letters of the alphabet, and you will be asked to judge whether these letters appear more often in the first or in the third position, and to estimate the ratio of the frequency with which they appear in these positions.

A typical problem read as follows:

Consider the letter R.
Is R more likely to appear in
__the first position?
__the third position? (check one)
My estimate for the ratio of these two values is ______:1.

Results of this study found a significant difference between the two positions: 69% of respondents judged the first position to be more likely for a majority of the letters while only 31% judged the third position to be more likely. Further, four other letters (K, L, N, V) were also judged by a majority of subjects to be more frequent in the first than in the third position. For each of these five letters, the median estimated ratio was 2:1. These results were obtained despite that fact that all five of these letters occur more frequently in the third letter position (Tversky & Kahneman, 1982a).

As with the representativeness heuristic, there are a number of variations in factors affecting availability. How quickly memory is searched and instances are retrieved or generated can depend on aspects of the stimulus such as its imaginability (an event which is easier to imagine is estimated to occur more frequently) or on aspects of the search process and its indexing via strength of association (strongly associated events are perceived as more frequent than events less strongly associated).

Anchoring and Adjustment

The last heuristic to be considered briefly here is that of anchoring and adjustment (Kahneman, Slovic, & Tversky, 1982). According to this heuristic, estimates are based on initial values and the estimation process consists of making adjustments from this anchor point. The adjustments being often insufficient, they lead to biased judgements. Tversky and Kahneman

(1974a) demonstrated this bias with two groups of high school students. Students were given five seconds to estimate a numerical expression that was written on the blackboard. One group estimated the product of 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1, while another group estimated the product of 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8. As predicted, the first group made larger estimates (median = 2,250) than the second (median = 512) with both groups being quite wrong (correct answer = 40,320). Tversky & Kahneman concluded that the difference in the two estimates is a result of the starting point of the series which the participants appear to have used as their anchor and the failure to sufficiently adjust their estimates from their starting points. This heuristic process is what Tversky and Kahneman believed was the reason for subjects non-normative judgements.

Another form taken by the anchoring and adjustment heuristic is one involving the evaluation of conjunctive and disjunctive events. This is the tendency for people to overestimate the probability of conjunctive events and underestimate disjunctive events. An example of a conjunctive event is commercial product development. For a product to be developed, it must pass a series of successive events. Each step may have a sizeable likelihood to succeed but the probability of reaching the overall goal remains low especially when the number of steps is large. According to the normative theory - in this case the multiplication rule - the outcome or goal is equal to the product of each successive step (Tversky & Kahneman, 1974). Conversely, in disjunctive events, where the probability of each step is low but the overall probability may be large if there are many steps, people tend to underestimate the probability of failure (e.g., the malfunction of a nuclear reactor or some other complex mechanism, situations where normative theory infers that the outcome is equal to the sum of each successive step). Thus in these complex situations, judgements do not match normative expectations.

Researchers from the heuristics and biases program (Bar-Hillel, 1980; Bar-Hillel, 1983; Bar-Hillel, 1990; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1973; Tversky & Kahneman, 1974; Tversky & Kahneman, 1982c) have argued that anchoring and lack of adjustment is one of the root causes of base-rate neglect in the cab problem. As already mentioned, the normative judgement to this problem (.41) differs markedly from the median empirical responses of the participants (.80). The empirical judgements correspond with the credibility of the witness, suggesting that the testimony is the anchor, and seem to be relatively unaffected by the relative frequency of Blue and Green cabs (little if any adjustment using the base rate).

There is a definite post-hoc flavour to this interpretation of why judgements are biased. To say that people anchor on the witness testimony is merely restating what was observed and does not explain why the testimony is chosen as the anchor. Why not choose the base rate? To say that the adjustment is insufficient reveals nothing about how adjustment unfolds and why it is insufficient. Thus it is clear that cognitive psychology must understand how anchors are set and how adjustments are made if non-normative judgements are to be fully understood. As of now, a theory of heuristics is still needed to remove their post-hoc character. An alternative viewpoint is to ignore the heuristic concept and look for an explanation elsewhere. Both approaches have been pursued to some degree. Let us look at some of each line of research.

Attempts to Understand the Neglect of Base Rates

Relevance and Causality

In studies of the cab problem conducted by Tversky and Kahneman, (1982c), base-rate information was utilised in the absence of individuating data. That is, when information pertaining to the witness' credibility was omitted from the problem, most participants responded with the base rate (either 15 or 85). It was also observed that base-rate information could exert control over the participants' expectations about the evidence. Here, Tversky and Kahneman replaced the sentence "a witness identified the cab as Blue" by "a witness identified the colour of the cab." Participants were then asked, "What is the probability that the witness identified the cab as Blue?" The median and modal response to this question was 15 while the correct answer is

[(.2)(.85) + (.8)(.15)] x 100 = 29. (Tversky & Kahneman, 1982b, p. 157) concluded that "in the absence of other data, ... the base rate was properly used to predict the target outcome and improperly to predict the witness' report".

From the data of these early studies, investigators such as Ajzen (1977) and Tversky and Kahneman (1980) have interpreted the neglect of base-rate information in terms of a causality heuristic, wherein the impact of the information depends on whether the judge perceives it as causally linked to the event in question. The intuitive judge would tend to ignore the population frequencies of engineers and lawyers because he or she cannot think of these frequencies as causing an individual to become an engineer or a lawyer (Tversky & Kahneman, 1982c). In contrast, the rich individuating information about the personality of the individual can be readily interpreted in terms of intuitive scripts or schemata that causally relate personality traits to occupations.

Ajzen (1977) and Tversky and Kahneman (1980) have further shown that when base rates have causal meaning, they tend to be more incorporated into probability judgements. Zuckerman (1978) found similar results and concluded that people make use of behavioural base rates when "they are able to form a meaningful script that accounts for this behaviour, and thus use this script in predicting the target's behaviour" (p.165). For example, if national statistics indicate that one out of every three people who drink and drive end up in an automobile accident then it would be quite justified to predict that if one's friends were to drive home while under the influence of alcohol, they too would have a one in three chance of being involved in an automobile accident (Zuckerman, 1978).

In relation to the cab problem, a different pattern of results was observed when the incidental base rate of cabs in the classic problem was replaced by a causal base rate of accidents (Tversky & Kahneman, 1980, p. 157). This was accomplished by replacing item (a) with (a'): "Although the two companies are roughly equal in size, 85% of cab accidents in the city involve

Green cabs and 15% involve Blue cabs". With this new problem, base rates were no longer ignored, although answers were highly variable. The median was 60 - halfway between the credibility of the witness (80) and the normative response (41). Tversky and Kahneman (1982b) argue that the base rates provided in (a') are causal because the difference in rates of accidents between companies of equal size readily elicit the inference that drivers of the Green cabs are more reckless and / or less competent than the drivers of the Blue cabs - implying that Green cabs are more likely to be involved in accidents. On the other hand, base rates provided in formulation (a) are believed to be incidental since the number of Blue and Green cabs in the city do not directly justify a causal inference making Green cabs more likely to be involved in accidents.

Bar-Hillel (1980) took this analysis one step further, proposing that the causal implications of base rates permit specific, individual characteristics, to be inferred from population characteristics. She argues that it is specificity rather than causality that determines whether or not base rates are utilised. Therefore, when asked to make judgements about a member or an item of a subset of a population (e.g., asking whether an individual is an engineer or a lawyer after being given information about the sample collected), specificity can be thought of as information pertaining to a subset of the population and seen as more relevant than information pertaining to the entire population. In these instances, the more specific information normatively supersedes the more general information (Bar-Hillel, 1980). She found that when base-rate information is not specific to the individual case then it is perceived as irrelevant in to judgements. Therefore, according to Bar-Hillel (1980), the "base-rate fallacy" occurs when the judgement task involves individuating information as well as base rates that are not concrete, causally meaningful, and specific to the judged individual case (Bar-Hillel, 1980).

Diagnosticity and Nondiagnosticity

Other explanations have been offered for the neglect of base rates in probability judgements. As already described, Kahneman and Tversky (1973) proposed that people use the representativeness heuristic to solve many types of probability problems. They believe that people judge the probability of an individual being an engineer rather than a lawyer according to the degree to which the personality description is more representative of or similar to the stereotype of an engineer than to the stereotype of a lawyer. Since base-rate frequencies do not bear upon the perceived representativeness of a particular individual, base rates are ignored in the presence of singular individuating information. Nisbett and Borgida (1975) and Nisbett, Borgida, Crandall, and Reed (1976) have demonstrated that the neglect of base-rate frequencies extends to a variety of social judgement tasks such as behavioural predictions, trait inferences, and causal attributions. These authors' explanation was that base rates, like other kinds of statistical data, are "remote, pallid, and abstract" whereas individuating information is "vivid, salient, and concrete" (Nisbett, Borgida, Crandall, & Reed, 1976, p. 24). Therefore, individuating information may serve as a trigger (e.g., the process of priming or availability) for a person's stereotype of a target's group membership.

The fact that people frequently disregard category-based information (base rates) in favour of individuating information is not surprising given that individuating information is clearly diagnostic of an individual's standing on some particular dimension. Hilton and Fein (1989) provide the following examples. If a man were attempting to determine how assertive a woman was and saw her shout someone else down, it makes sense for him to ignore his beliefs about the passivity of women in general and to focus instead on her individual behaviour. Similarly, if this man observed another man cowering in the corner at some formal gathering, it makes sense for him to rely on the man's behaviour rather than on his general beliefs about the assertiveness of men. In both instances, the individuating information is clearly more relevant to his judgement and is more diagnostic of the target's standing on the dimension in question than is the information available from knowing the person's category membership (Locksley, Borgida, Brekke, & Hepburn, 1980).

Diagnostic individuating information is not the only information that has been shown to influence individual probability judgements. Nondiagnostic individuating information (i.e., information believed to be neutral and having little or no relevance to the judgement task) has also been shown to influence the judgements of others. In one study (Locksley, Hepburn, & Ortiz, 1982), participants rated target individuals on traits that were stereotypically characteristic of either diurnals (day people) or nocturnals (night people). When participants received only information about whether a target was diurnal or nocturnal, their judgements of the targets were consistent with their beliefs about the categories. Diurnals were seen as dependable and self-controlled, whereas nocturnals were seen as more unpredictable and rebellious. However, when participants received information about whether a target was diurnal or nocturnal in combination with individuating information that was nondiagnostic for the judgement task (e.g., information indicating that the target's mother was a nurse), participant's beliefs about diurnals and nocturnals no longer had as strong an effect on their judgements of the targets.

Correspondingly, Nisbett, Zukier, and Lemley (1981) found that when participants were presented with two target individuals, one identified as an engineer and the other identified as a music major, participants predicted that the engineer would withstand higher intensities of shock. Again, when participants were provided with additional information that had previously been judged as nondiagnostic for shock tolerance (e.g., information indicating that the targets were Catholic), the difference in predicted shock tolerance was attenuated. These findings led both Locksley et al. (1982) and Nisbett et al. (1981) to conclude that social stereotypes appear to exert relatively little impact on judgements of individuals about whom even a minimal amount of nondiagnostic information is available. An alternative way of looking at these results is that participants used otherwise uninformative information in a way that made their judgements more conservative than if this information had not been presented.

Overall, these attempts at explaining base rate neglect have had some degree of success in that situations were identified where base rates would or would not be taken into consideration (e.g. causal versus non causal or specific versus non-specific contexts). As well, situations were also identified where information such as base rates would be unnecessarily used. But there remains something less than satisfying about the approach. Base rates are said to be neglected because they are not seen as causal. This is tantamount to explaining the insufficient adjustment heuristic by having recourse to another heuristic controlling causality perception. The strong reliance on the witness testimony is said to occur because it is more representative of the kind of evidence that the court would normally use. This is explaining the anchoring heuristic by recourse to the representativeness heuristic. When the explanations are attempting to go beyond mere reformulation of what was observed, there is a risk of infinite regression and/or of circularity in the chain of explanations in this framework. For this reason alone, it is worthwhile considering different theoretical approaches altogether.

Two Alternative Theoretical Frameworks

Frequentist Format

A summary of the long standing thinking on the matter of understanding and explaining probabilistic reasoning is given in a nutshell by Kahneman and Tversky (1973):

> In making predictions and adjustments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgements and sometimes lead to severe systematic errors (p. 237).

In opposition to this, Gigerenzer and colleagues (1991a; Gigerenzer & Hoffrage, 1995; Gigerenzer & Murray, 1987) argue that such an approach of identifying errors and explaining them has little to do with probability and statistics reasoning. Gigerenzer's main argument against this program is that even after 25 years of research, still very little is known about the processes underlying human inference, Bayesian or otherwise. He asserts that the representativeness heuristic used to explain the neglect of base rates was only loosely defined then (in the early 1970s) and still remains "a vague and ill-defined notion" (Gigerenzer & Hoffrage, 1995; Gigerenzer & Murray, 1987). He writes that the normative theory of probability invoked by the research program is "a very narrow kind of neo-Bayesian view that is shared by some theoretical economists and cognitive psychologists... and is not shared by proponents of the frequentist view of probability that dominates today's statistics departments..." (Gigerenzer, 1991b, pp. 86-87; 1993a; 1993b; 1994; 1996; Gigerenzer & Hoffrage, 1995; Gigerenzer & Murray, 1987).

The frequentist position is based on two related areas of research: a) the work of Hasher and Zacks (1979; 1984) on the automatic encoding of frequency information, and b) the evolutionary perspective of Cosmides and Tooby (1996). Hasher and Zacks have shown that humans are very sensitive to frequency information about a variety of stimuli such as letters, numbers, words etc. and process this frequency information automatically, without effort. Cosmides and Tooby argue that humans in their natural habitat experience information in series of events rather than as probabilities or percentages; because of this, they would have evolved an information representation structure that is frequentist (Gigerenzer & Hoffrage, 1995). Providing evidence that many animal species, including humans, are highly sensitive to changes in frequency distributions in their environments, the evolutionary perspective argues that frequentist cognitive algorithms, acquired through natural sampling, were developed to handle the manner in which the frequency is most often experienced. Based on this, Gigerenzer suggests that asking whether the mind is Bayesian or not is not at the heart of the matter. Rather, how information is represented in the mind and how best to present information in probabilistic reasoning tasks to take advantage of the mind's format are the fundamental questions needing to be addressed. The mind's format, he argues, is frequentist in nature and, in order to detect the mind's cognitive algorithm, the information provided in probabilistic problems such as the cab problem must match this representational format.

In order to evaluate their claim about the theorised mind's format, Gigerenzer and his colleagues have compared answers to conditional probability problems presented in various formats. For instance, Gigerenzer and Hoffrage (1995) used four different formats: 1) a standard probability expressed in percentages; 2) a standard frequency, which provided the same amount of information as the standard probability but expressed the numerical quantities as relative frequencies; 3) a short probability, which only provided the necessary information expressed in percentages required to solve the problem; and 4) a short frequency that only provided the required information in relative frequencies. Here are two illustrations of the two frequency formats used by Gigerenzer & Hoffrage (1995).

Cab Problem - Standard Frequency Format

> 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*) 15 out of every 100 cabs in the city are Blue cabs.
>
> (*ii*) A witness identified the cab as a blue cab. The court tested the witness' ability to identify cabs under the appropriate visibility conditions.
> 12 out of every 15 Blue cabs are identified as blue cabs.
> 17 out of every 85 Green cabs are also identified as blue cabs.
>
> Here is a new representative sample of cabs that are identified by the witness as blue cabs. How many of these cabs do you expect to actually be Blue? _____ out of _____

Cab Problem - Short Frequency Format

> 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*) 29 out of every 100 cabs in the city are identified by a witness as being blue cabs.
>
> (*ii*) A witness identified the cab as a blue cab. The court tested the witness' ability to identify cabs under the appropriate visibility conditions.
> 12 out of every 100 cabs in the city are Blue cabs *and* identified by a witness as a blue cab.
>
> Here is a new representative sample of cabs that are identified by the witness as blue cabs. How many of these cabs do you expect to actually be Blue? _____ out of _____

Overall, Gigerenzer's team found that presenting the information in a frequency format, especially the short version, elicited the most normative responses. It should be noted however that the standard and short frequency formats are dramatically different, conceptually and structurally, from the original cab problem posed by Tversky and Kahneman (1982c). It is not just that numbers have been changed from percentages to integers. By changing the probabilistic information to discrete units (presenting the relative frequencies followed by the question, "so many *out of* so many"), there is a reasonable possibility that the problem was not perceived as the same as the classical scenario. It is also possible that the frequentist formulation actually leads the way to more relativity in the solution. Either way, Gigerenzer's team may not have been studying the same cognitive process that Tversky and Kahneman intended to. This is not trivial since it greatly impacts on what interpretations can be made by comparing the two lines of evidence. If the underlying cognitive processes are different, then comparisons of which type of research produced the most normative answers does not inform us on the nature of the process but simply on the relative amounts of bias in two different activities. Within the frequentist viewpoint however there is a clear expectation of normative answers.

Neither the Bayesian nor the frequentist model provide much of a theoretical framework to study the extent to which problem formulation alone can influence judgement. To provide such a framework, it is possible to turn to the *Principles of Conversational Conventions.*

Principles of Conversational Conventions

According to conversational inference (Hilton, 1995b; Schwarz, Strack, Hilton, & Naderer, 1991), judgements must be considered in the context in which the information is presented. Using Grice's (1975) conversational conventions, Hilton (1991; 1995a; 1995b) proposes that many of the biases found in psychological research may be less attributable to heuristic factors such as representativeness and anchoring and adjustment than to inferences about the social context of the problem, which in turn is guided by conversational assumptions.

Grice (1975) proposes a theory of conversational inference based on the Principle of Co-operation, which is assumed to underlie conversation, and the conversational maxims derived from it. Briefly, the Co-operative Principle assumes that a speaker accepts the purpose or direction of the talking exchange in which he or she is engaged, and makes his or her contribution as required, at the stage at which it should occur. This is achieved by following four maxims, each implying certain kinds of attributes about the interlocutor. In the cab problem, these are attributes about the witness (Hilton, 1995b). The four conversational maxims are: a) the maxim of quality, b) of manner, c) of quantity, and d) of relation.

The maxim of quality concerns the truth value of the information provided by a speaker (here, the witness). If the hearer (i.e., participant who makes a probability judgement) attributes properties such as sincerity, honesty, reliability, and competence to the speaker, then the hearer may well consider the truth value of an utterance to be high. In the case of the cab problem, the attribute of reliability of the witness statement would be relied on very heavily to arrive at an answer because sincerity, honesty and competence are assumed high on the part of the witness, in which case the reliability level becomes highly diagnostic. Thus, from a conversational analysis, it is theoretically understandable why participants would want to make probability judgements that match the witness' testimony, i.e. why they would anchor on the information provided by the witness.

According to the maxim of manner, the speaker is enjoined to provide information that is brief, orderly, clear, and unambiguous. The assumption is that people do not want long-winded, incomprehensible explanations. Hilton (1995b) concludes that experimenters may not notice ambiguities in their questions that are systematically reinterpreted by participants, which in turn lead to systematic biases in the results obtained. Accordingly, the information provided in the cab problem should suffer a lesser risk of misinterpretation if it follows these attributes. However, it is not clear that the classic formulation does, whereas the versions of the problems in Experiment

3 and the frequentist problems in Experiment 4 may fare better.

The maxim of quantity pertains to the perceived information value of an utterance. During a conversation, speakers should not burden hearers with information they are already likely to know. The content of the conversation should be as informative as required for the current purposes exchange but not overly or repetitively informative (Hilton, 1990; 1995b). According to this maxim, people should use the base rate information in solving the cab problem because the very presence of it would make the participants assume that it is not superfluous. Clearly, that is not the case. Note however that people also know that not all speakers fully respect all of these maxims and can also adjust their understanding accordingly. Overall then, this maxim does not help very much in trying to account for probabilistic judgements. A similar criticism may be applied to the fourth maxim, that of relation.

The maxim of relation implies that the information given is intended to be relevant to the goals of the interaction. Those listening to a conversation are entitled to assume that any relevant information they are not likely to know has been included, and that the information that has been included is relevant. Otherwise why mention it? On the face of it, this maxim would also predict that base rate information should be used. On the other hand, if people have already anchored on the witness testimony and, additionally, are not fully trusting the problem formulation to respect the maxim of quantity, then they may infer that base rate has little relation to the solution of the problem and ignore it. It could also happen that base rate is deemed irrelevant for lack of causal relevance as mentioned before, or simply because the participants tried yet failed to detect the relevance. In any of these alternatives, the prediction is one of insufficient adjustment of the anchor (witness statement) as a function of base rate, if indeed such is the cognitive process in solving the cab problem. Thus, like the maxim of quantity, this maxim can easily make opposing predictions, which severely limits its usefulness.

The witness reliability is the only attribute quantified explicitly in the cab problem. Thus participants would be left to assume that the witness is simultaneously very sincere, honest and competent. On the other hand, reliability is the only attribute that might matter for the base rate information yet no quantity is provided for it. In addition, as stated above, it is dubious that the maxim of manner is very well respected in the classic cab problem formulation because the percentage data tend to blur what is a relative to what. Thus, applying these maxims to the cab problem, and assuming a desire to co-operate from the participant, a conversational analysis predicts an over-reliance on the witness information and under-utilisation of base rates (Hilton, 1995b).

Objectives of the Thesis

The existence of base rate neglect has been empirically demonstrated many times. However, the extent of this neglect is variable and we lack a comprehensive theory of why the neglect occurs and the extent to which it should be expected in any specific situation. Several specific theoretical accounts have been proposed (e.g., Relevance and Causality, Diagnosticity and Non-diagnosticity, Frequentist Format, and Conversational Conventions), but more research is still needed to determine where and when these theoretical constructs apply, if at all. Specifically, the explanatory power of the frequentist format and of conversational conventions are tested. This is one objective of the thesis. Another purpose of the thesis is to propose a novel method of presenting frequency information and to distinguish between conditions that foster more normative answers in probabilistic reasoning and conditions that do not. The problem that serves as the medium for the various experimental manipulations is always the cab problem.

Hilton (1995b) reports that the phenomenon of underusing base rates is "restricted to 'word problems' in which the base rate information is presented verbally to participants in the form '30% of the group are engineers'"(p. 255). On the other hand, studies that have presented base rates in the form of on-line learning trials have shown that participants are more likely to use base-rate information (e.g., Christensen-Szalanski & Beach, 1982; Medin & Bettger, 1991; Medin & Edelson, 1988). This is consistent with the observation that on-line studies of contingency judgements, which are also probabilistic in nature, have reported that people can normally assess contingencies quite well (Baker, Berbrier, & Vallée-Tourangeau, 1989; Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Shanks & Dickinson, 1987). On-line procedures have the potential of proving very useful in comparing the frequentist and the heuristic theoretical stances because they allow for the direct provision of trial-by-trial frequency data while preserving the exact wording of the classic, off-line, probability question. This is what Experiment 1 sets out to achieve.

Experiment 1

To format the cab problem on-line, the information must be presented in a series of trials. For this purpose, a graphical computer display was created where Green and Blue coloured cabs crossed the screen and were involved in accidents. A witness provided testimony as to what he[1] believed to be the colour of the cab. Based on the frequentist approach (e.g., Cosmides & Tooby, 1996; Gigerenzer, 1991b; Gigerenzer & Hoffrage, 1995), it was expected that participants who acquired the information on-line would more closely approximate normative judgements. In addition, the on-line format lent itself to a solution based on the direct calculation of the conditional probability of the cab colour given the witness' testimony [P(B|b)], a computation which is much simpler than Bayes' Theorem. This is achieved by conceptualising and arranging the trial-by-trial information into a 2 x 2 contingency table (see Table 1). Cell a corresponds to the frequency of the witness indicating that the cab was blue when the cab was truly Blue; cell b is the frequency of the witness indicating the cab to be blue when it was truly Green; cell c is the frequency with which the witness said green when the cab was truly Blue; and cell d is the

---

[1] Reference to the witness will be in the masculine gender as the witness presented in the on-line experiments was always a man.

frequency of the witness stating green when the cab was truly Green. Therefore, the answer to the question, "What is the probability that the cab was Blue given that the witness said it was blue" is a / (a + b). This answer corresponds to the answer also obtained by way of Bayes' Theorem[2]. Although research participants are never instructed to use such a contingency table[3], research in contingency judgement (Baker, Berbrier, & Vallée-Tourangeau, 1989; Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993) has shown that people can readily estimate such conditional probabilities[4]. Thus, this experiment offers participants an opportunity to provide more normative judgements by providing frequentist information and by enabling the possibility of simpler calculations. By comparison, judgements made in an off-line format, which is included in Experiment 1 as a replication of the classic finding, are expected to produce the previously demonstrated neglect of base rates.

Table 1: A 2 x 2 contingency table of the relative frequencies of the Base Rates and Individuating Information in the Cab Problem.

		True Cab Colour	
		Blue	Green
Witness	blue	a	b
Testimony	green	c	d

In the off-line condition, participants were presented with probabilistic information in the form of word problems identical to those described by Tversky and Kahneman (1982c) except that they were posted as static text on a computer screen. The question posed in these problems

2 For instance, if the cell frequencies of the classical cab problem where the base rate of Blue and Green cabs were 15% and 85% respectively and the reliability of the witness being correct was 80%, then the Bayesian answer would be 41.4% (Equation 3). Out of 40 trials, the corresponding frequencies would be 12, 17, 3, and 68 for cells a, b, c, and d respectively. Thus cells a / (a + b) works out to be 12 / (12 + 17) = 12/29 = 41.4% which is the same as the Bayesian solution.

3 The contingency table is only used as a device by the experimenter to clarify the design of the experiment.

4 The research in question does not claim that people compute a/(a+b) to arrive at contingency judgements but it is

was: "What is the probability that the cab involved in an accident was Blue given that a witness said it was blue?"

In the on-line condition, participants saw series of 40 cabs having accidents on the computer screen and a witness providing statements as to the possible colour identification in the proportions required by the desired probability to be tested. At the end of 40-trial[5] series, a black car crossed the screen, the witness made a last visually unverifiable colour statement, and the participants answered the question: "What is the probability that the *last* cab involved in an accident was Blue given that a witness said it was blue?" All participants completed both types of problem format thus allowing for a within-subject comparison of on-line versus off-line problem format. Further, each format contained 9 problems: low, medium and high base rate conditions that were completely crossed with low, medium and high reliability conditions, and counterbalanced for the target colour of the cab.

Method

Participants. Twenty-four undergraduate students from the University of Ottawa participated in this experiment. During recruitment, students were informed that each participant would have an equal chance (1 in 24) of winning a prize worth $50. All participants in the experiments reported in this dissertation research were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 1992).

Apparatus. An IBM compatible PC computer with a SVGA colour monitor (resolution was set to 640 X 480 pixels) was used to display the stimuli and record the data. The program was written in Microsoft Visual Basic.

---

sufficient for the purpose of Experiment 1 that participants here could do so.

[5] The choice of 40 for the number of trials is based on the research of Baker, Mercier, Vallée-Tourangeau, Frank, & Pan (1993), Dickinson, Shanks, & Evenden (1984) and Shanks (1985).

Procedure. Stimulus examples for the on-line (graphics) portion of the experiment are depicted in Figures 2 and 3. The scenario consisted of a taxi cab, either Blue or Green depending on the trial, travelling along a road then suddenly veering off and crashing into a house. At the crash site, a red paint symbol was displayed to depict that an accident had taken place.

After 40 trials, a 41^{st} cab, black with a big yellow question mark painted on it, appeared and an eye-witness provided information as to what he believed to be the colour of the cab involved in

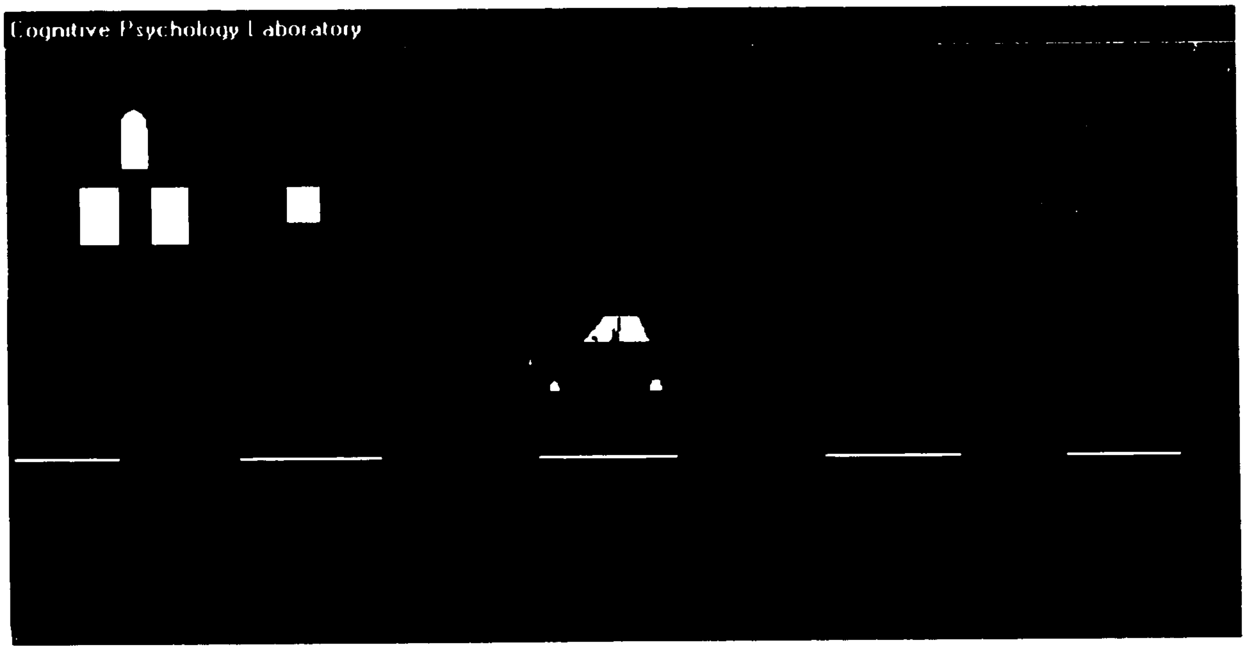

Figure 1. Example of a Blue cab trial in the on-line format.

the accident, indicating that the cab was either blue or green (see Figure 1). The witness was a 2.5 cm by 1.25 cm figure standing in the top centre of the screen with a cartoon-like speech bubble indicating what he believed to be the colour of the cab. At the time of the accident, a 5 cm by 3.5 cm figure of a house situated in the top left of the screen became covered in red (see Figure 2). The cab was a 2.5 cm by 1.5 cm figure that moved across the screen from right to left at the rate of 5 centimetres per second, and a 8 cm by 6.5 cm dialogue box that would appear once the

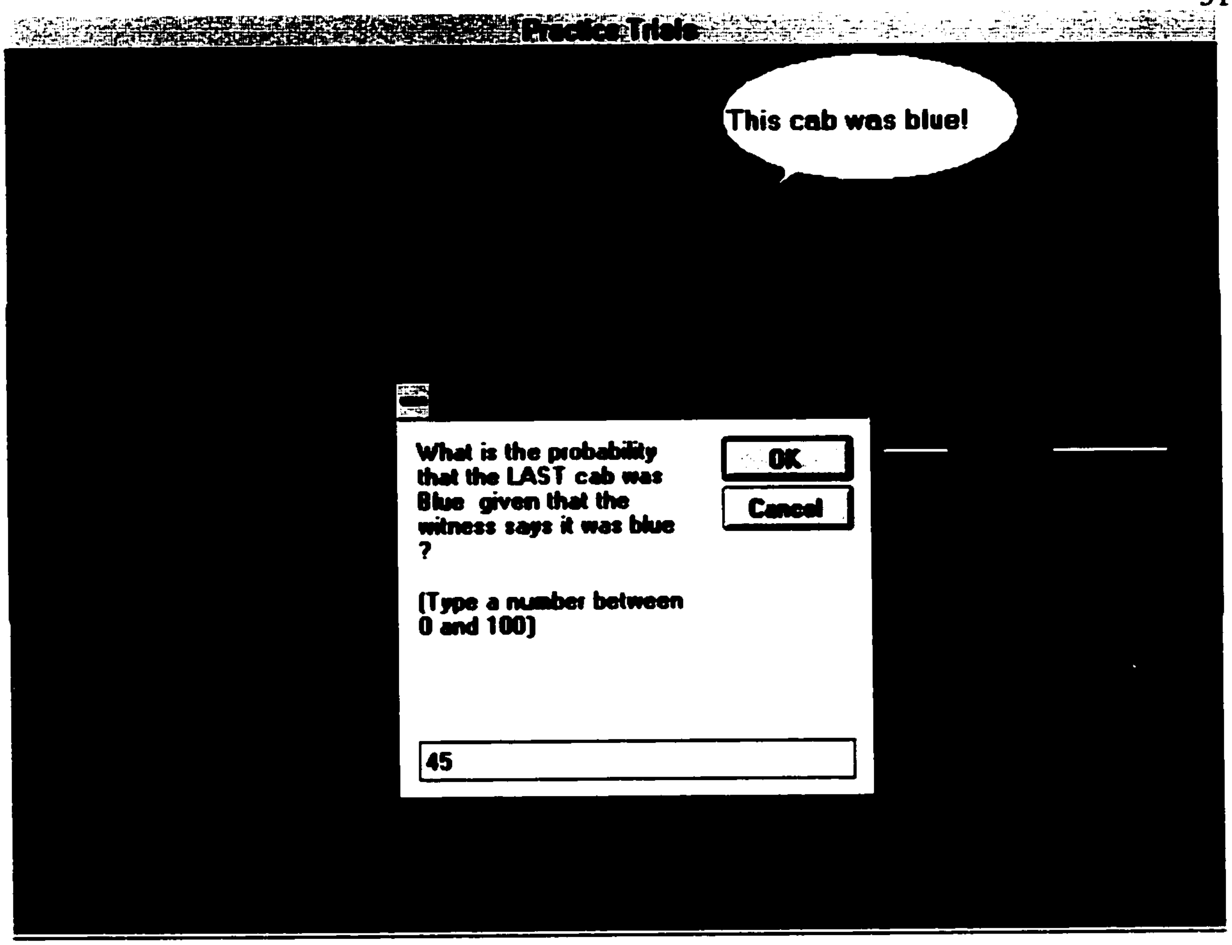

Figure 2: Example of the witness statement, the accident, and the question.

unknown cab had passed, allowing participants to make their judgements. Participants made judgements for 9 separate games, each series of 40 trials representing one game

The off-line (word) problems are depicted in Figure 3. Participants made judgements for 9-word problems, one problem representing each level of base rate crossed with each level of witness reliability. The format of the word problems was very closely modelled after Bar-Hillel (1980) and Tversky & Kahneman (1982c).

The design of this experiment was completely within-subjects and involved 3 levels of base rates (15, 50, and 85), 3 levels of witness reliability (referred hereafter as individuating information: 20, 50, and 80), and 2 levels of problem format (on-line and off-line). Thus, participants made judgements for 18 experimental problems. In addition, two practice problems, one on-line and one off-line, were also provided to all participants but not included in the analysis

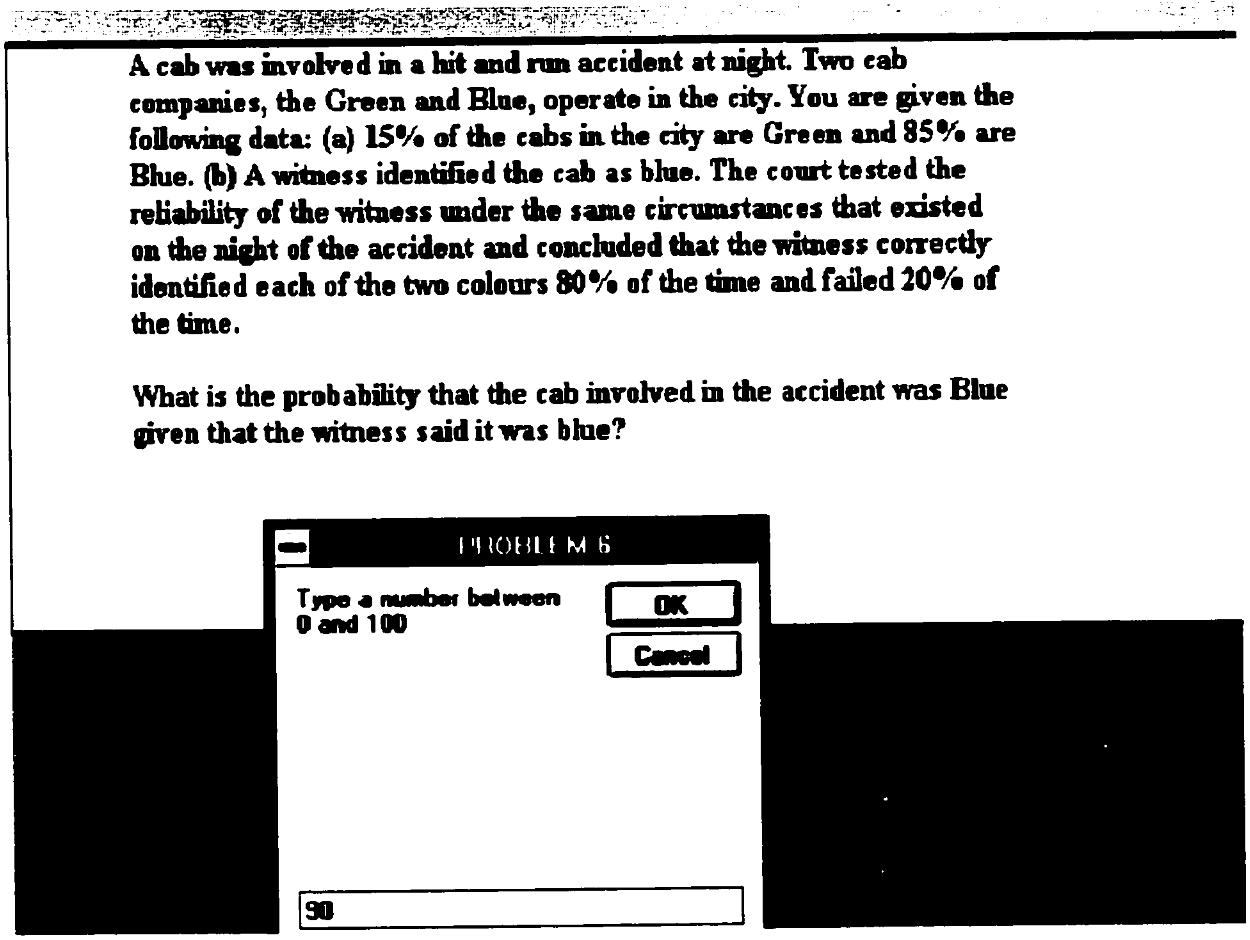

Figure 3: Sample word problem.

of the experiment. Between each problem was a change in both base rate and individuating information. To control for the order of problems, 24 unique series (one series for each participant) consisting of the 18 experimental problems (9 on-line and 9 off-line problems) were created. In presenting the series of twenty problems, the first two problems were always the two practice problems (the on-line followed by the offline). The 18 experimental problems were then presented: the first three problems in the on-line format followed by the first three problem in the off-line format, then the next three problems in the on-line format followed by the off-line format. This was repeated until all 18 problems had been presented.

The order of presentation of the series was devised by arranging the problems in a 3 by 3 table in a manner similar to that in Table 2. Then each problem was placed in the first position only once. Determining the second and third problems in the series was based on moving one problem down and to the right. The fourth problem in the series was determined by moving one down from the first problem and the fifth and sixth problems determined by following the movement procedure just described (moving one problem down and to the right). The same procedure for determining problems 4 to 6 was followed for problems 7 to 9. Thus, in a block of 9 series, each problem was positioned in each of the 9 positions only once. For example, one series consisted of nine problems presented in the following order: practice problem, LBLI, MBMI, HBHI, LBMI, MBHI, HBLI, LBHI, MBLI, and HBMI. The next series in this block of 9 would be as follows: practice, MBMI, HBHI, LBLI, MBHI, HBLI, LBMI, MBLI, HBMI, and LBHI. Determining the order of the second block of 9 series was based on the following movement pattern: one problem left and one problem down (e.g., practice, LBLI, HBMI, MBHI, LBMI, HBHI, MBLI, LBHI, HBLI, and MBMI). The last block of problems consisting of 6 series was determined by moving one problem up and one problem right (e.g., practice, LBLI, MBHI, HBMI, MBLI, HBHI, LBMI, HBLI, LBHI, and MBMI). The order of presentation for the on-line problems consisted of following the sequence order from left to right, whereas, the off-line format consisted of the sequence order from right to left. This created a unique series of problems for each problem format for every participant.

Cab colour was counterbalanced according to the target colour indicated in the question. Half the problems asked for the probability of a Blue cab given that the witness reported it to be blue while the remaining problems asked about a Green cab given the witness reported green. For even numbered participants, problems 1 to 5 asked about Blue cabs and problems 6 to 9 asked about Green cabs. For odd numbered participants, problems 1 to 5 asked about Green cabs and problems 6 to 9 about Blue cabs. This manipulation ensured that participants paid attention to all the information presented in order to properly solve any given problem.

Table 2: Design of Experiment 1 as a function of base rate and individuating information.

		Individuating Information		
		.20	.50	.80
	.15	LBLI	LBMI	LBHI
Base rate	.50	MBLI	MBMI	MBHI
	.85	HBLI	HBMI	HBHI

L, M, and H correspond to low, medium and high base rates (B) and individuating (I) information

Participants sat in front of the computer and read the page of instructions presented on the screen. The exact instructions were as follows:

> We are conducting research to better understand how people make decisions under situations of uncertainty.
>
> To do this, we will present you with a series of events where Blue and Green cabs are involved in accidents at night. Each accident will be seen by a witness. When all the information has been presented, a witness will indicate what he believes to be the colour of the last cab. You will not have seen the colour of that last cab and your job will be to tell us what are the chances that the witness is correct. You will make judgements for 9 separate "games", each game involving an assessment of the witness' testimony. Each game will involve a new and independent witness.

Next, participants were presented with a practice session involving 40-trials presented in random order: 10 trials of a Blue cab involved in an accident and the witness reporting the cab to be blue; 10 Blue cabs and the witness reporting green; 10 Green cabs and the witness reporting blue; and 10 Green cabs and the witness reporting green. Following these, a black coloured cab with a yellow question mark appeared on the screen, was involved in an accident, and the witness reported what he believed to be the colour of this unknown cab. Then, the question shown in Figure 2 would appear on the screen: "What is the probability that the LAST cab was Blue (or Green) given that the witness says it was blue (or green)? (Type a number between 0 and 100)" [colours in parentheses refer to the counterbalanced scheme and did not appear on the screen].

Upon entering an answer to the practice on-line problem, participants were presented with a practice off-line problem (e.g., Figure 3) that contained the following information: the base-rate information in part (a) consisted of 25% Green cabs and 75% Blue cabs; and the individuating information in part (b) consisted of the witness correctly identifying the two colour cabs 100% of the time and failing 0% of the time. Upon completion of the two practice problems, participants had the opportunity to ask questions prior to proceeding to the 18 experimental problems. The actual frequency of trials for the 9 on-line problems are presented in Table 3.

Statistical analysis. A three-way repeated measures analysis of variance (ANOVA) was performed on the data. In this and the following experiments, .05 was selected as the criterion for tests of significance. Univariate F values are reported throughout and all within-subject effects were corroborated with multivariate tests. For each experiment, the observed experimental data is graphically compared to normative estimates.

Table 3: Frequencies of Blue and Green Cabs involved in accidents and witness reliability for on-line problems in Experiment 1[6].

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	1	27
green	5	7

Problem 1: LBLI
P(B|b) = 1/28 = .04

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	4	16
green	16	4

Problem 2: MBLI
P(B|b) = 4/20 = .20

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	7	5
green	27	1

Problem 3: HBLI
P(B|b) = 7/12 = .58

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	3	17
green	3	17

Problem 4: LBMI
P(B|b) = 3/20 = .15

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	10	10
green	10	10

Problem 5: MBMI
P(B|b) = 10/20 = .50

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	17	3
green	17	3

Problem 6: HBMI
P(B|b) = 17/20 = .85

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	5	7
green	1	27

Problem 7: LBHI
P(B|b) = 5/12 = .42

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	16	4
green	4	16

Problem 8: MBHI
P(B|b) = 16/20 = .80

Witness Testimony	True Cab Colour: Blue	True Cab Colour: Green
blue	27	1
green	7	5

Problem 9: HBHI
P(B|b) = 27/28 = .96

---

[6] Normative judgements [e.g., P(B|b)] were the same for each problem (e.g., LBLI) in both the on-line and off-line formats.

Results and Discussion

Figure 4 depicts participant's overall mean probability judgements. Mean judgements for the on-line and off-line formats are plotted for each of the nine problems along with the corresponding normative answers. The pattern of results indicate that mean judgements increased as the level of individuating information increased (judgements increasing from panel to panel) and increased as the level of base rates increased (judgements increasing within each panel). Also, there appear to be little if any differences between the on-line and off-line problem formats.

The contention that judgements were most sensitive to the influence of the witness' testimony is supported by the main effect of Individuating Information, F (2, 46) = 218.22, MSE = 394.55, with a very large effect size (η^2 = .905). Base rates also influenced judgements, but to a

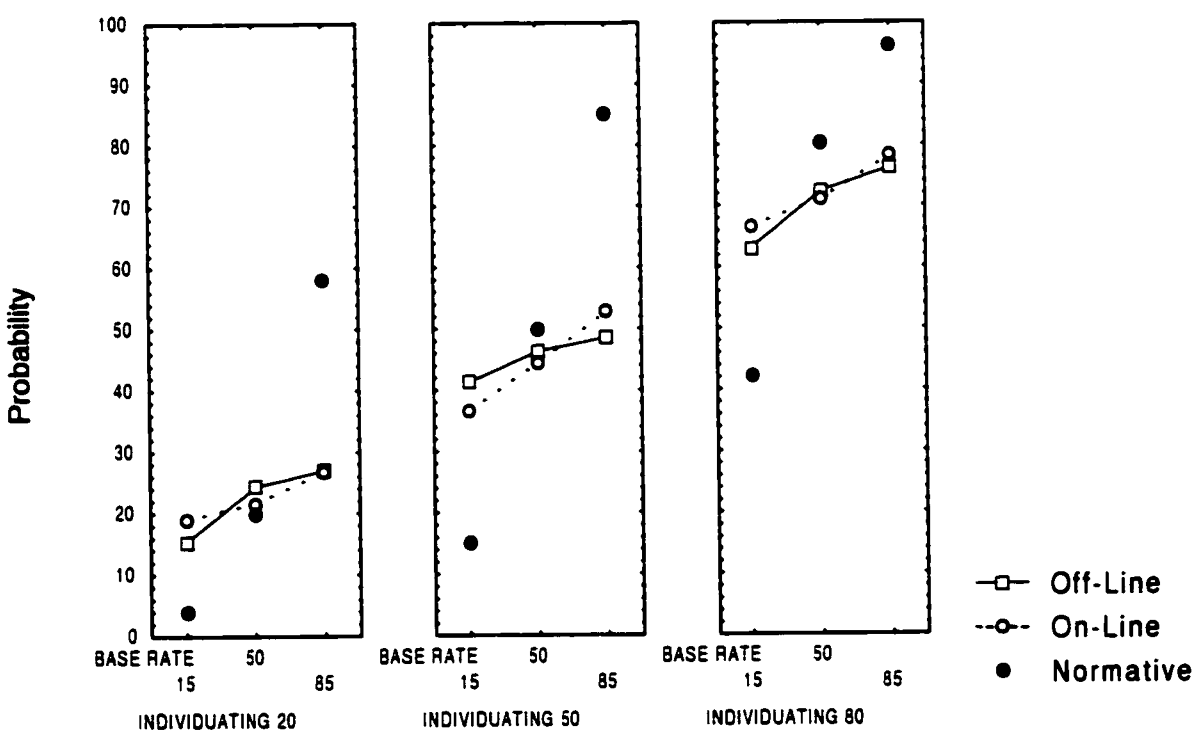

Figure 4: Mean probability judgements as a function of BASE RATE by INDIVIDUATING INFORMATION by CONDITION in Experiment 1.

lesser degree than the witness' testimony. This is confirmed by the main effect for Base Rates, $F(2, 46)=16.08$, $MSE=287.23$, $\eta^2 = .411$. These effects held true irrespective of whether the information was presented on-line or off-line. Neither a higher order interaction of any kind, nor a main effect for problem format (on-line vs. off-line) was found to be significant.

These results add to our understanding of probabilistic reasoning in two major ways. First, the lack of differences between problem formats suggest that the frequentist format of trial-by-trial presentation was not enough to produce more normative responding than the word-problem format. This is contrary to what was expected based on a frequentist analysis (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995)

Second, though the effect for the base rate information is small compared to the individuating information, it is not small in an absolute sense but in fact very large ($\eta^2 = .411$) (Cohen, 1969; Kirk, 1995). Further, Figure 4 shows that within each panel of Individuating Information, the lines slope quite a bit as a function of changes in the base rate information. This occurs for both the on-line and off-line conditions. This result is not consistent with the notion of a base-rate fallacy (Bar-Hillel, 1980; 1983; 1990; Kahneman & Tversky, 1973; Tversky & Kahneman, 1982c).

Histograms of the raw probability judgements (see Figure 5 and Figure 6) corroborate the strong influence of the witness' testimony on the judgements of most participants. In both the off-line and especially the on-line conditions, the most frequently observed response corresponds to the witness' reliability (20, 50, and 80). This pattern is evident irrespective of the level of base rate or the format of problem presentation.

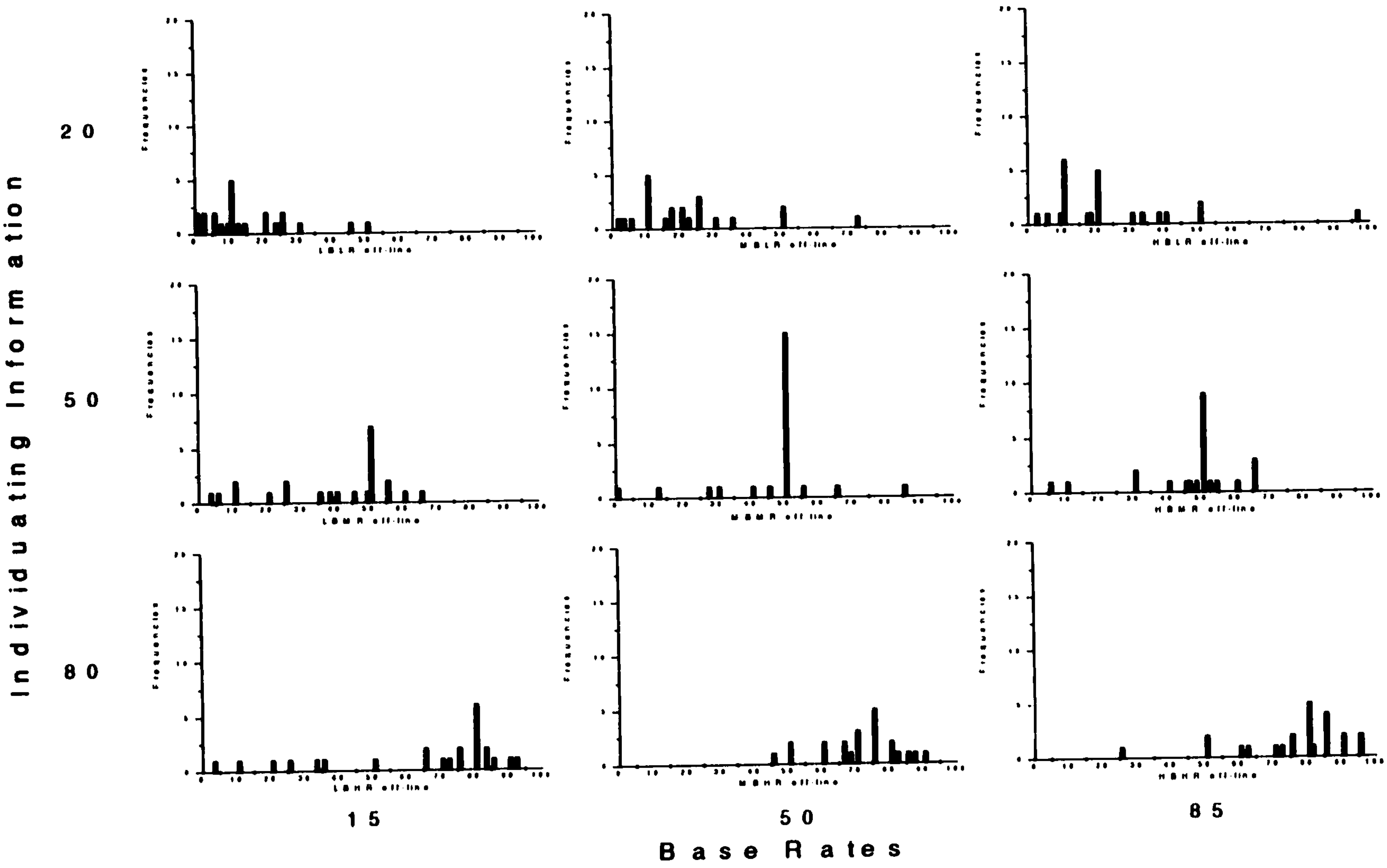

Figure 5: Histogram for the raw probability judgements made in the off-line condition of Experiment 1.

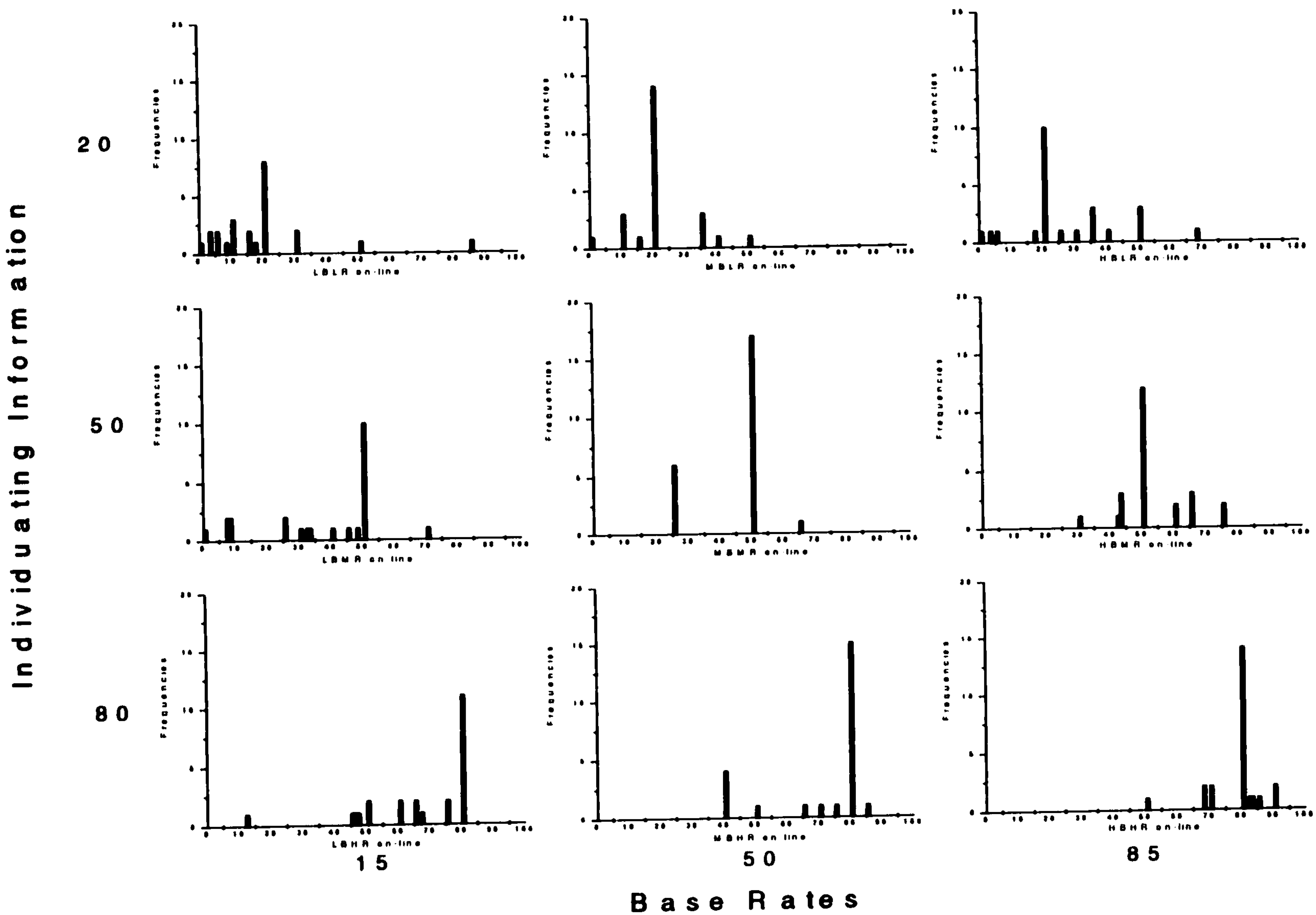

Figure 6: Histogram of the raw probability judgements made for the on-line condition in Experiment 1.

This pattern is very distinctive even in the on-line format which is surprising since the frequency format of this condition was expected to induce more normative judgements and less reliance on the witness information. In any event, the results of this experiment suggest that the mere presentation of frequency information in a trial-by-trial format is not enough to reduce the bias of the individuating information.

The almost identical judgement patterns of mean probability judgements produced by the on-line and the off-line formats in this experiment rules out the possibility that past research failed to obtain normative answers because the calculation of Bayes' Theorem or a similar formulae was too difficult for participants. In this experiment, participants could have solved the problems simply by counting the number of times the target cab colour appeared and the witness correctly identified that colour and dividing by the number of times the witness identified cabs as the target colour (solving the problems by way of cells a / (a + b)). Yet they failed to do so in spite of independent evidence (Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993) indicating that participants are capable of such a task.

The lack of differences between the two formats suggests that there is something beyond mere frequencies influencing judgements. A double mechanism might have caused the current results to occur. With a low base rate (15) and high witness reliability (80), the mean judgement made here was 62 in the off-line format and 66 in the on-line format (the normative response being 41). Variations of the "classic" cab problem have found that when the incidental base rates are made causal (Tversky & Kahneman, 1980), specific and relevant (Bar-Hillel, 1980), then the median response is approximately 60 - lying somewhere between the reliability of the witness (80) and the normative response (41). Perhaps because participants were given both the on-line and off-line problem in a within-subjects design, with the off-line always following the on-line problems, the on-line format may have somehow induced participants to view the base rate as being specific and causally relevant to the judgement task. Or perhaps the exposure to both formats simply drew attention to all the components of the question. That is, the on-line format graphically and explicitly illustrated the relationship between the base rates (cab colours) and the individuating information (the witness' testimony). In turn, the activation of this judgement strategy may have been carried over to the off-line problems in a mechanism of prior activation (Ginossar & Trope, 1987). The next two experiments investigate the plausibility of this double mechanism to account for why there was a demonstrable sensitivity to base rates in both on-line and off-line tasks in Experiment 1.

Experiment 2

The objective of Experiment 2 was to investigate whether enhancing relevance and causality of the base rate information in combination with an on-line format would yield more normative judgements. A variation of the cab problem was developed where cabs could have accidents only some of the time and the probability of a given colour was assessed conditional on the cab's accidental status. Recall that in Experiment 1 all cabs were involved in an accident on each trial. In Experiment 2, Blue and Green cabs are sometimes involved in accidents, sometimes not.

This modification of the task was implemented for two reasons. One, it makes the on-line problem even more similar to the contingency judgement tasks in which participants must take into account two conditional probabilities simultaneously: that of the effect given the cause (i.e., probability of a Blue cab given that it was involved in an accident) and that of the effect in the absence of the cause (i.e., probability of a Green cab given it was <u>not</u> involved in an accident). As in Experiment 1, since contingency judgements have been shown to be quite accurate under many circumstances (Kao & Wasserman, 1993; Wasserman, Dorner, & Kao, 1990), making the revised version of the cab problem more analogous to contingency judgements was expected to result in more normative judgements. The more normative answers could arise either from the participants computing a / (a + b) or from an associative process such as that hypothesised in the contingency judgement area (Baker, Berbrier, & Vallée-Tourangeau, 1989; Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Shanks & Dickinson, 1987; Wasserman, Dorner, & Kao, 1990).

The second reason for the modified task was that the occurrence of accidents on less than all the trials made it possible to consider the accident itself as a correlate of cab colour. Since the problem deals with the colour of the cab involved in an accident (or not involved in an accident in the case of the Green cabs), using the accidental status of the cab as a condition for the evaluation of a conditional problem should maximise the relevance of its rate of occurrence.

The information was presented trial-by-trial in three on-line conditions: 1- the standard formulation with instructions only about the witness [P(B|W)]; 2- standard formulation with a witness not explicitly mentioned, but present, and the conditional statement based on the cab's accidental status [P(B|A)W]; and 3- a modified formulation without any witness and the conditional based on the cab's accidental status [P(B|A)].

Therefore, if the on-line format truly induces a consideration of the causal relevance of the base rate information, then the P(B|W) condition should produce a pattern of results similar to Experiment 1 where base rates were used reliably but where judgements are more influenced by the witness' testimony and far from normative expectations. The P(B|A) and, to some extent, the [P(B|A)W] conditions are expected to yield judgements closer to the normative answer.

Method

Participants and Apparatus. Sixty undergraduate students from the University of Ottawa participated in this experiment. During recruitment, students were informed that each participant would have an equal chance of winning $50. The apparatus was the same as that of Experiment 1.

Design and Procedure. The design of this experiment was a 3 x 3 x 3 mixed model. The three levels of base rate and individuating information (witness reliability and/or accidental status) were the same as in Experiment 1 and constituted two within-subjects factors. The three levels of the between-subject factor were conditions P(B|W), P(B|A)W, and P(B|A) with 20 participants randomly assigned to each condition. Condition P(B|W) was the same as the on-line problem format in Experiment 1 with the exception that Green and Blue cabs were sometimes involved in accidents and other times passing safely. The frequencies of accidents and non-accidents for Blue and Green cabs were identical to the frequencies of the witness statement of cab colour, as detailed in Experiment 1. The witness reliability and accidental status had a 1:1 correspondence. Therefore, for the conditions P(B|W) and P(B|A)W, every cab that was involved in an accident also had the witness indicating the cab to be blue, regardless of its true cab colour. Similarly, cabs not involved in accidents were always reported by the witness to be green. This resulted in the same normative judgement for each of the three conditions.

An example of the question in condition P(B|W) was "What is the probability that the LAST cab was Blue given that the witness said it was blue?" For condition P(B|A)W, the question was "What is the probability that the LAST cab was Blue given that it was involved in an accident." In this condition, a witness is present and makes statements, but a normative solution require only two pieces of information: the Blue and Green cabs that are involved in accidents. The witness information is superfluous. Finally, no witness was present in condition P(B|A), and the question asked was the same as for P(B|A)W. The sequencing of the series of trials and the counterbalancing of cab colour was controlled for according to the procedures detailed in Experiment 1.

Results and Discussion

The pattern of judgements obtained is depicted in Figure 7. The results are similar to those of Experiment 1. Judgements increased with each increase in individuating information (compare across panels) and with each increase in base rate (compare within panels). These contentions are supported by main effects for Individuating Information, $F(2, 114) = 67.71$, *MSE*

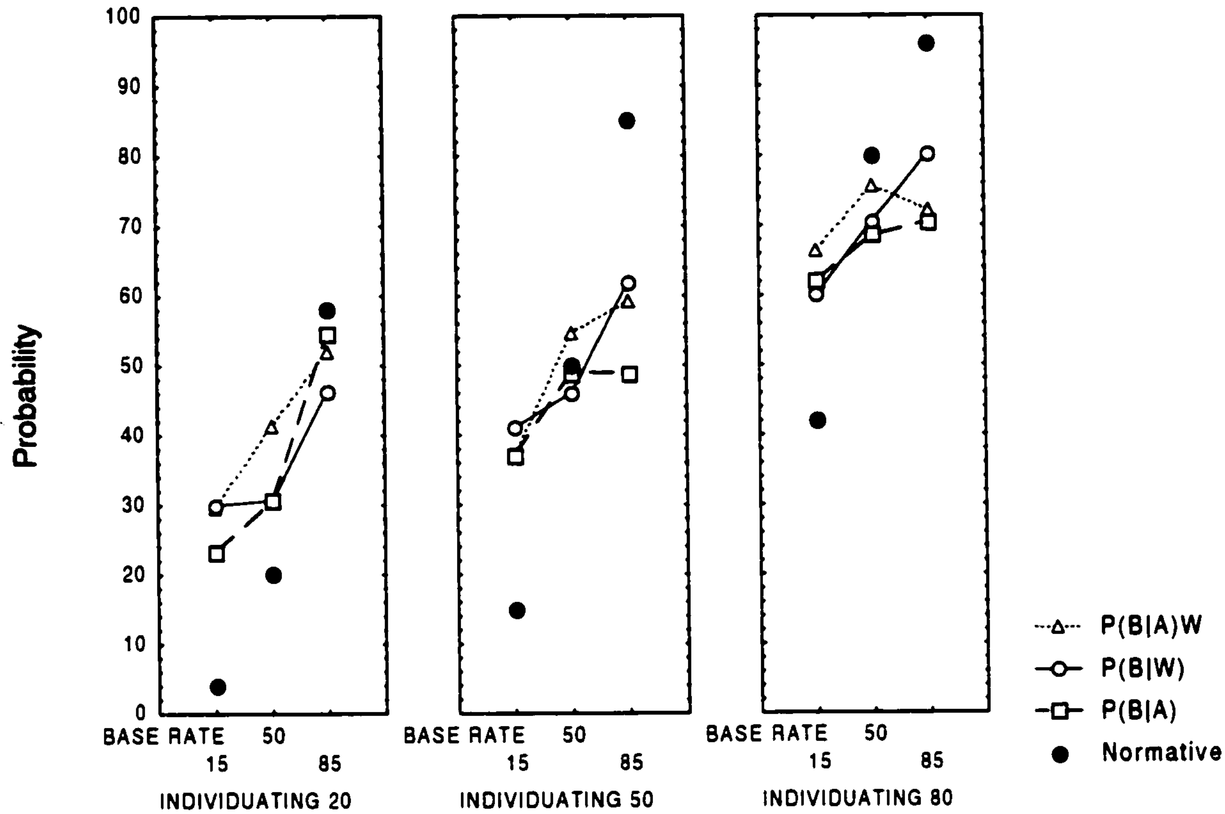

Figure 7: Mean Judgements as a function of BASE RATE by INDIVIDUATING INFORMATION by CONDITION in Experiment 2.

= 703.37, η^2 = .543 and for Base Rates, $F(2, 114) = 25.39$, $MSE = 557.38$, $\eta^2 = .308$. No other significant main effects or higher order interactions were found.

The histograms in Figure 8, Figure 9, and Figure 10 show that individual judgement patterns to be very similar across the three groups with distributions that are disperse and variable. What appears to be common in all conditions is a modal response corresponding to the Individuating Information (especially when the Base Rates are 50) apparently combined with the Base Rate, perhaps in multiplicative fashion.

The fact that neither of the two conditions involving the accidental status generated more normative judgements fails to support the hypothesis of relevance and specificity. At the same time, in both Experiments 1 and 2, participants demonstrated a clear sensitivity to base rates as well as to the individuating information. This begs the question: How do participants arrive at judgement computationally? If they calculate the conditional probability using cells a / (a + b) then their answers would be less sensitive to the individuating information and would not require the base rate information at all. So they must be doing something else. In isolation, the judgement pattern from Experiment 2 could be accounted for by the simple heuristic of anchoring on the individuating information and adjustment relative to the changes in base rates. Yet, as stated in the general introduction, this heuristic explanation is unsatisfactory for a number of reasons. First, the heuristic principle falls short of specifying how much adjustment should be made. Second, the heuristic principle does not indicate why it is the individuating information that is used to anchor and the base rates to adjust, and not vice versa. If neither relevance and specificity, nor anchoring and adjustment, nor a frequentist format of presentation are the answer, then we must turn to a different theoretical approach to understand the psychological process underlying conditional probability judgement and its variations. This is what Experiment 3 sets out to do.

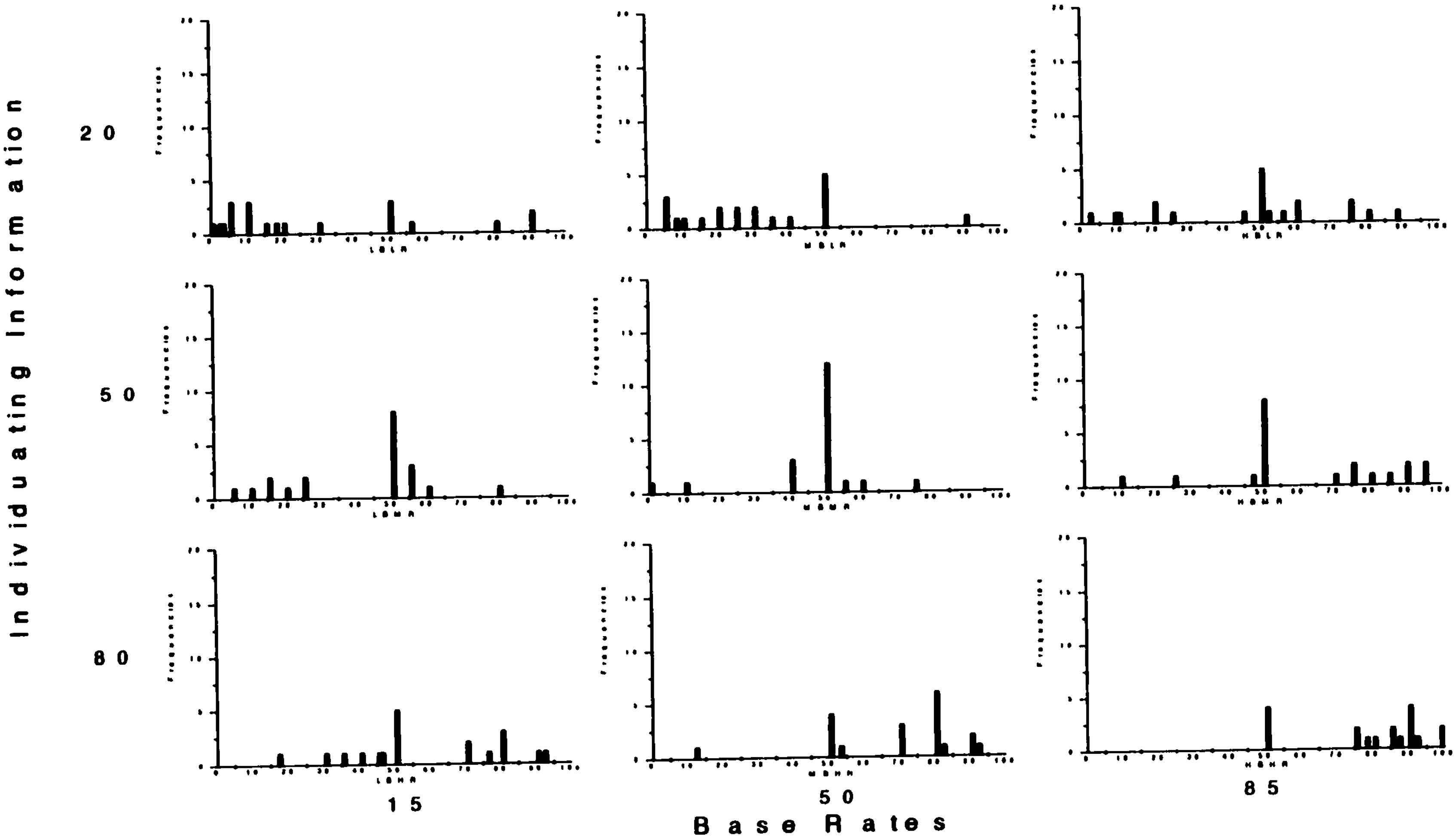

Figure 8: Histogram of the raw probability judgements for the P(B|W) condition in Experiment 2.

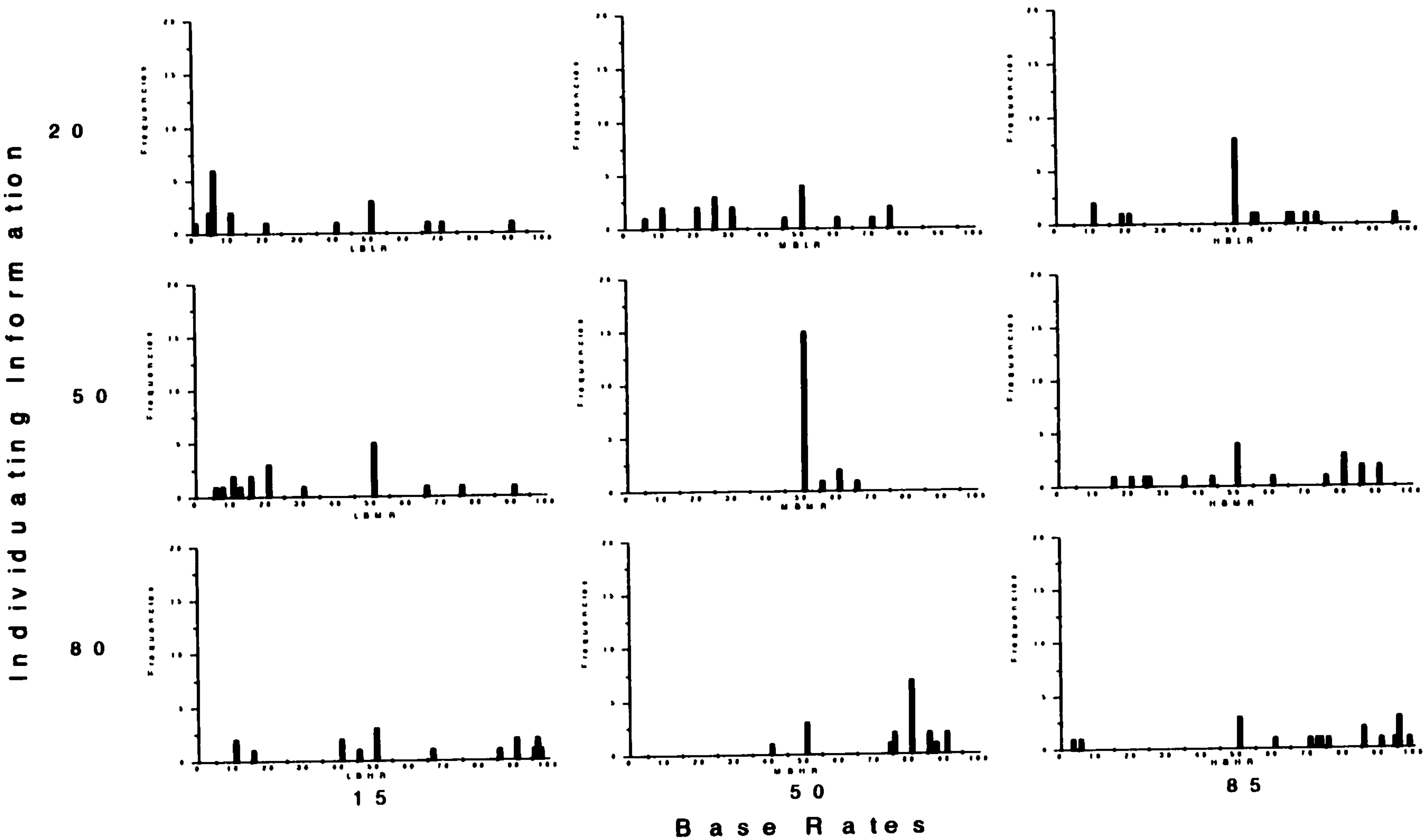

Figure 9: Histogram of the raw probability judgements for the P(B|A)W condition in Experiment 2.

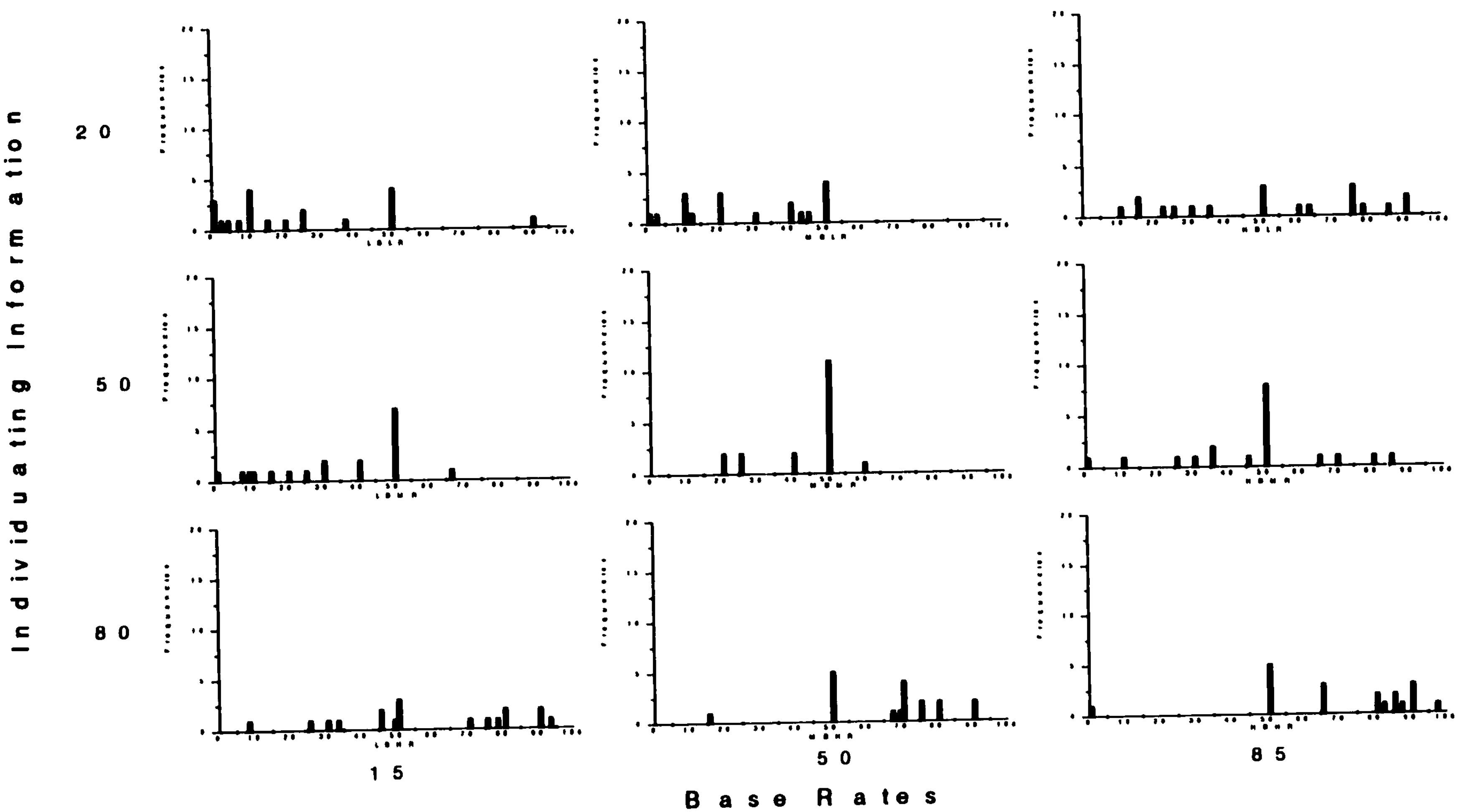

Figure 10: Histogram of the raw probability judgements for the P(B|A) condition in Experiment 2.

Experiment 3

In the general introduction, conversational analysis was described as an alternative theoretical framework in which to consider judgement under uncertainty. From the standpoint of conversational analysis, biased judgements in the cab problem would be influenced in large part by the witness testimony due to the maxim of quality (properties of sincerity, honesty, reliability, and competence) and by the way in which the problem is formulated due to the maxim of manner (property of ambiguity). Experiment 3 was devised to test these two maxims.

Four between-subjects conditions were constructed to address the maxim of quality, attributes of sincerity, honesty, reliability, and competence. Two conditions involved the witness as the conditional source of individuating information and two conditions involved the accidental status of the cab as the individuating source. This manipulation was retained in spite of its lack of detectable effect in Experiment 2 because a conversational analysis predicts that the accidental status as the conditional source should less readily elicit the notions of sincerity, honesty, and competence due its more mechanical, less human, nature as a source of information.

The other dimension of the design involved the maxim of manner, specifically the property of ambiguity. The information in the cab problem reports that the witness is correct 80% of the time, leaving it to the problem-solver to construct a mental formulation of this information. The correct formulation corresponds to 80% of the Blue cabs being identified as blue. It would be easy, however, for the problem-solver to misconstrue this as meaning 80% of reported blue cabs are truly Blue. These two proportions differ radically. In a contingency table such as the one depicted in Table 1, the first statistic corresponds to the proportion of cells a / (a + c) while the second corresponds to the proportion of cells a / (a + b). The latter proportion is not the information conveyed in the problem, but it is the correct answer sought. In order to derive the frequencies a and b from the written word problems, a normative judge would first need to calculate: cell a = P(Blue) · P(blue|Blue) and cell b = P(Green) · P(blue|Green), and then solve by taking a / (a + b). However, if problem solvers mistakenly think that the percentage correct given to them corresponds directly to a / (a + b) rather than a / (a + c) then it is no wonder they neglect base rates and give the percent correct as their judgement. Therefore, to test this hypothesis, two problem formulations were constructed: one that retains the ambiguity of the original studies and one that attempts to remove the ambiguity by clearly stating what (cells) the individuating information pertains to (see examples below).

The on/off-line manipulation was not retained as it had little if any effect in the first two experiments and the conversational analysis made no specific prediction concerning it.

Finally, to reduce the size of the design and the costs of printing individual questionnaires, the base rate manipulation dropped the intermediate value and retained only the two crucial points of low and high base rates.

Method

Participants. One hundred and fifty-five undergraduate students enrolled in first and second year Psychology courses at the University of Ottawa served as participants in this experiment. No financial remuneration was offered in this experiment.

Materials. Problems presented in the four conditions were again closely modelled after those of Tversky and Kahneman's (1980) cab problem. Examples of the four types of problems are presented in Table 4.

Table 4: Example problems in the four between-subjects conditions of Experiment 3

P(B|W) – Unambiguous A cab was involved in a hit and run accident at night. Two cab companies, the Blue and the Green, operate in the city. You are given the following data:(*i*) 15% of the cabs in the city are Blue and 85% are Green; (*ii*) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly declared that the cab was blue in 80% of the cases where the cab was really Blue and incorrectly declared that the cab was blue in 20% of the cases where the cab was really Green. What is the probability that the cab involved in the accident was Blue rather than Green? (Please respond with a number between 0 and 100).	**P(B|W) – Ambiguous** A cab was involved in a hit and run accident at night. Two cab companies, the Blue and the Green, operate in the city. You are given the following data: (*i*) 15% of the cabs in the city are Blue and 85% are Green; (*ii*) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time. What is the probability that the cab involved in the accident was Blue rather than Green? (Please respond with a number between 0 and 100).
P(B|A) – Unambiguous A cab was involved in a hit and run accident at night. Two cab companies, the Blue and the Green, operate in the city. You are given the following data: (*i*) 15% of the cabs in the city are Blue and 85% are Green; (*ii*) Based on the records from the cab companies data bases, it appears that cabs were involved in accidents in 20% of the cases where they were Blue and also involved in accidents in 80% of the cases where they were Green. What is the probability that the cab involved in the accident was Blue rather than Green? (Please respond with a number between 0 and 100).	**P(B|A) - Ambiguous** A cab was involved in a hit and run accident at night. Two cab companies, the Blue and the Green, operate in the city. You are given the following data: (*i*) 15% of the cabs in the city are Blue and 85% are Green; (*ii*) There is a possibility that the cab was Blue. The court tested the reliability of the relation between accidents and cab color under the same circumstances that existed on the night of the accident and concluded that knowing that a cab had an accident allowed to correctly identify each one of the two colors 80% of the time and fail 20% of the time. What is the probability that the cab involved in the accident was Blue rather than Green? (Please respond with a number between 0 and 100).

Design and Procedure. Two levels of base rate (high = 85 and low = 15) and three levels of individuating information (low = 20, medium = 50, and high = 80) were presented within-subjects. The four conditions P(B|W) Ambiguous, P(B|W) Unambiguous, P(B|A) Ambiguous, and P(B|A) Unambiguous were all presented between-subjects. This resulted in a 2 x 2 x 2 x 3 mixed factorial design.

Participants were run in large groups consisting of volunteers from one Introductory Psychology class and two second-year psychology classes. Each participant was given one-questionnaire booklet consisting of 6 word problems corresponding to one of the four between-subjects conditions. Participants worked on the problems individually at their own pace and submitted their booklets to the experimenter when completed. Sequencing and counterbalancing were implemented as in Experiments 1 and 2.

Results and Discussion

Mean probability judgements are depicted in Figure 11. Four aspects of the results stand out in this figure. First, mean judgements increased from panel to panel in relation to the corresponding increase in levels of the individuating information. This is supported statistically by a significant effect for the Individuating information factor; F (2, 280) = 150.98, MSE = 427.50, η^2 = .519. This replicates the results of Experiments 1 and 2 and, indeed, the results of Kahneman, Slovic and Tversky (1982).

Second, the average judgement increases within panel in relation with the corresponding changes in base rates, as is confirmed by a reliable main effect of base rate; F(1,140) = 68.66, MSE = 907.40, η^2 = .329. This also replicates Experiments 1 and 2 and is consistent with previous reports showing that base rates <u>can</u> be taken into consideration (Tversky & Kahneman, 1982c).

Third, judgements in the ambiguous condition were lower on average than judgements in the unambiguous conditions, as verified in the significant main effect of Ambiguity, F (1, 140) =

3.96, MSE = 906.39, η^2 = .028. This difference makes intuitive sense in that ambiguity should reduce confidence which, in turn, should lead to more conservative estimates (i.e. closer to zero).

These three main effects need to be qualified in view of the three significant two-way interactions of Witness/Accidental status by Individuating Information, F (2, 280) = 6.36, MSE = 427.50, η^2 = .043; Witness/Accidental status by Base Rates, F (1, 140) = 4.77, MSE = 907.40, η^2 = .033; and Ambiguity by Base Rate, F (1, 140) = 8.95, MSE = 907.40, η^2 = .06. However, even these interactions are best understood in the higher order light of the fourth phenomena shown in Figure 11, that of the notable difference between the judgements made in the P(B|A) Ambiguous condition versus the judgements made in all three others. On average, judgements made in the P(B|A) Ambiguous condition varied rather little from panel to panel in the figure, certainly less so than any of the other conditions. This indicates that the Individuating information had the least impact on the accidental status estimates in the ambiguous condition and is statistically supported by the significant three way interaction of Witness/Accidental status by Ambiguity by Individuating Information, F (2, 280) = 10.57, MSE = 427.50, η^2 = .07. Simultaneously, the steeper slope for the judgements in the P(B|A) Ambiguous condition within each panel of the figure suggests that base rates may have had more impact on these judgements than any other condition. However, the three way interaction of Witness/Accidental status by Ambiguity by Base rate did not reach significance (p = .075). Thus the lesser impact of Individuating information appears to be the main cause of the generally more normative judgements obtained in this condition.

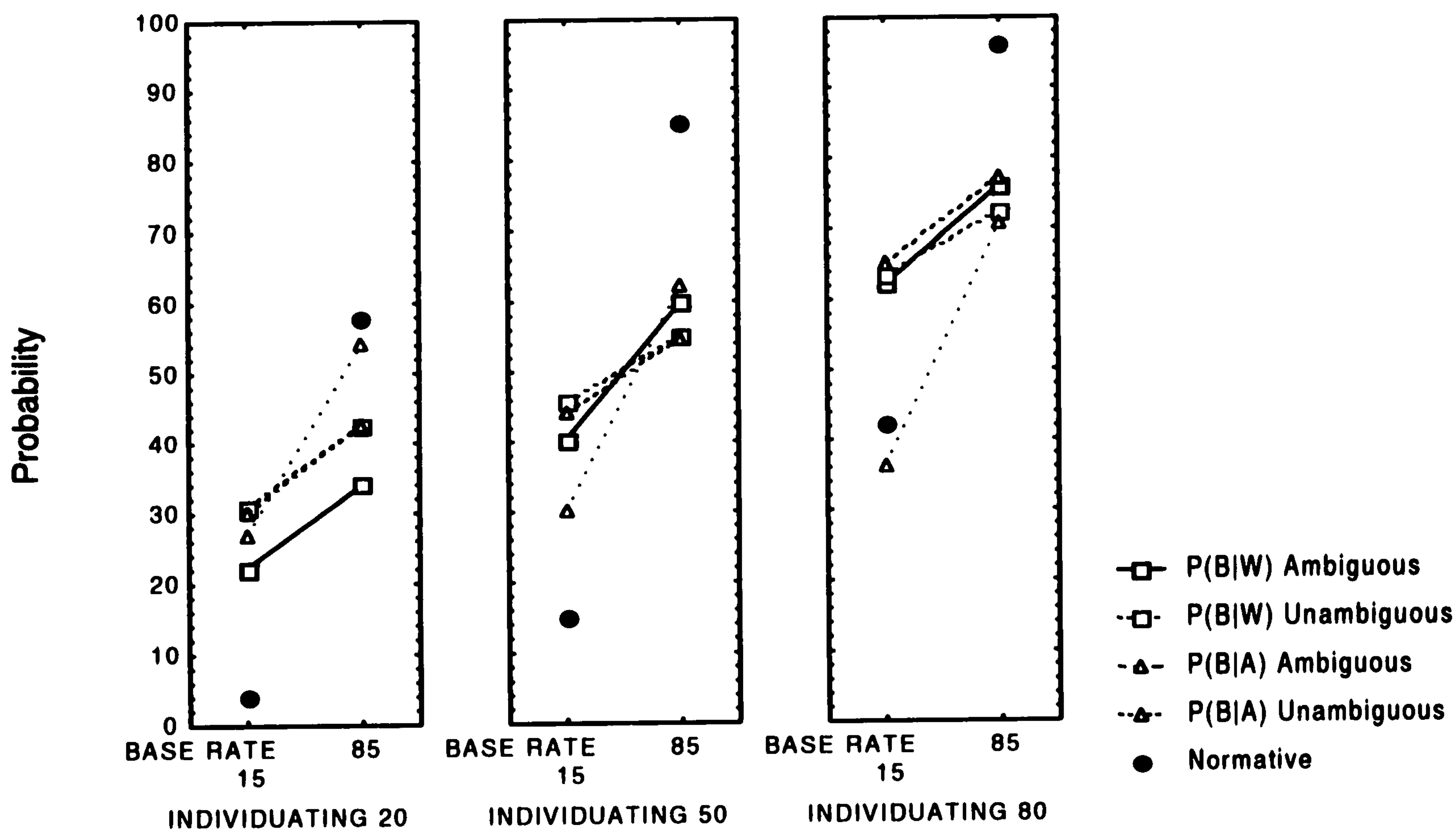

Figure 11: Mean probability judgements as a function of BASE RATES, INDIVIDUATING INFORMATION, and CONDITION in Experiment 3.

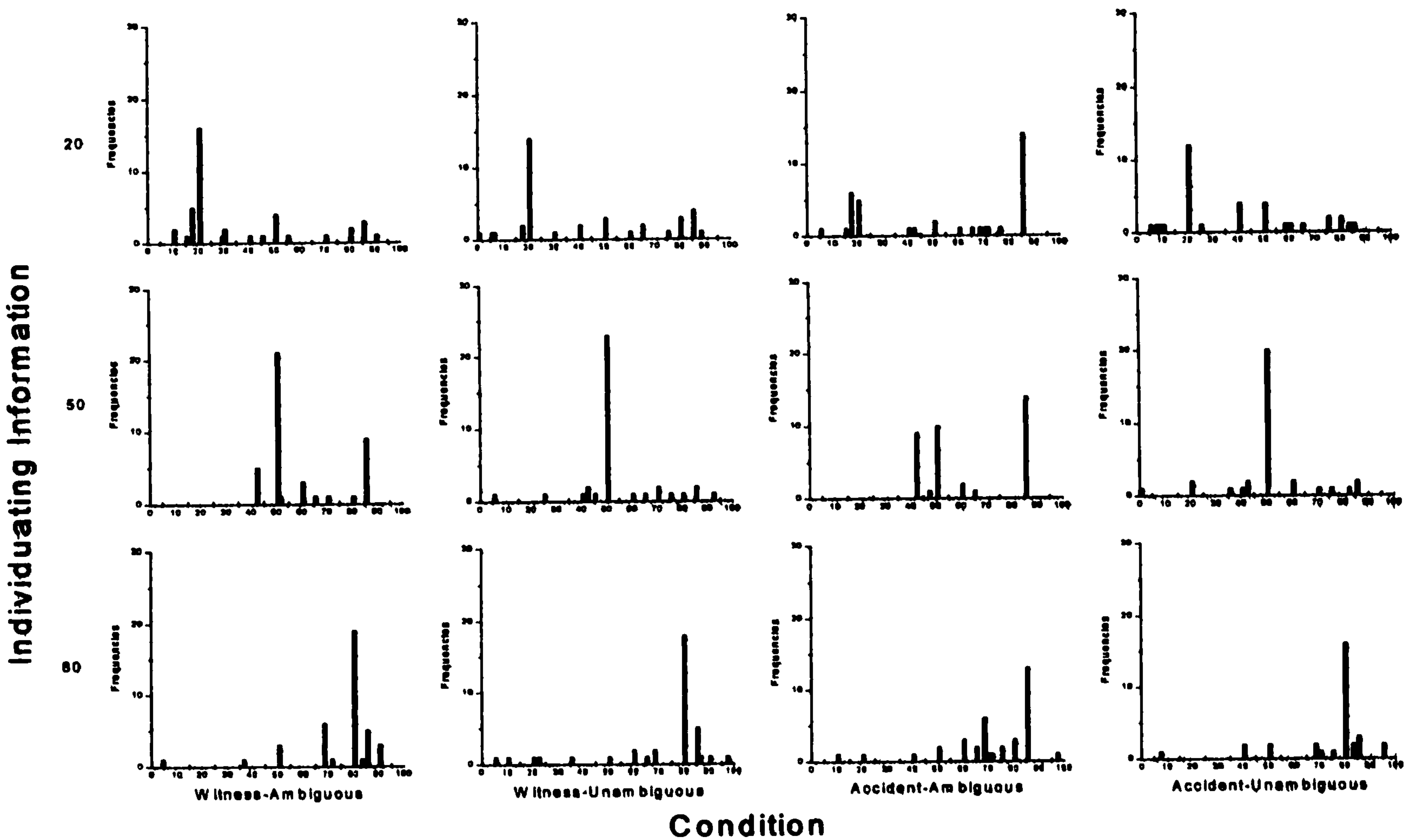

Figure 12: Histogram of raw probability judgements for the High Base Rate (85) conditions of Experiment 3.

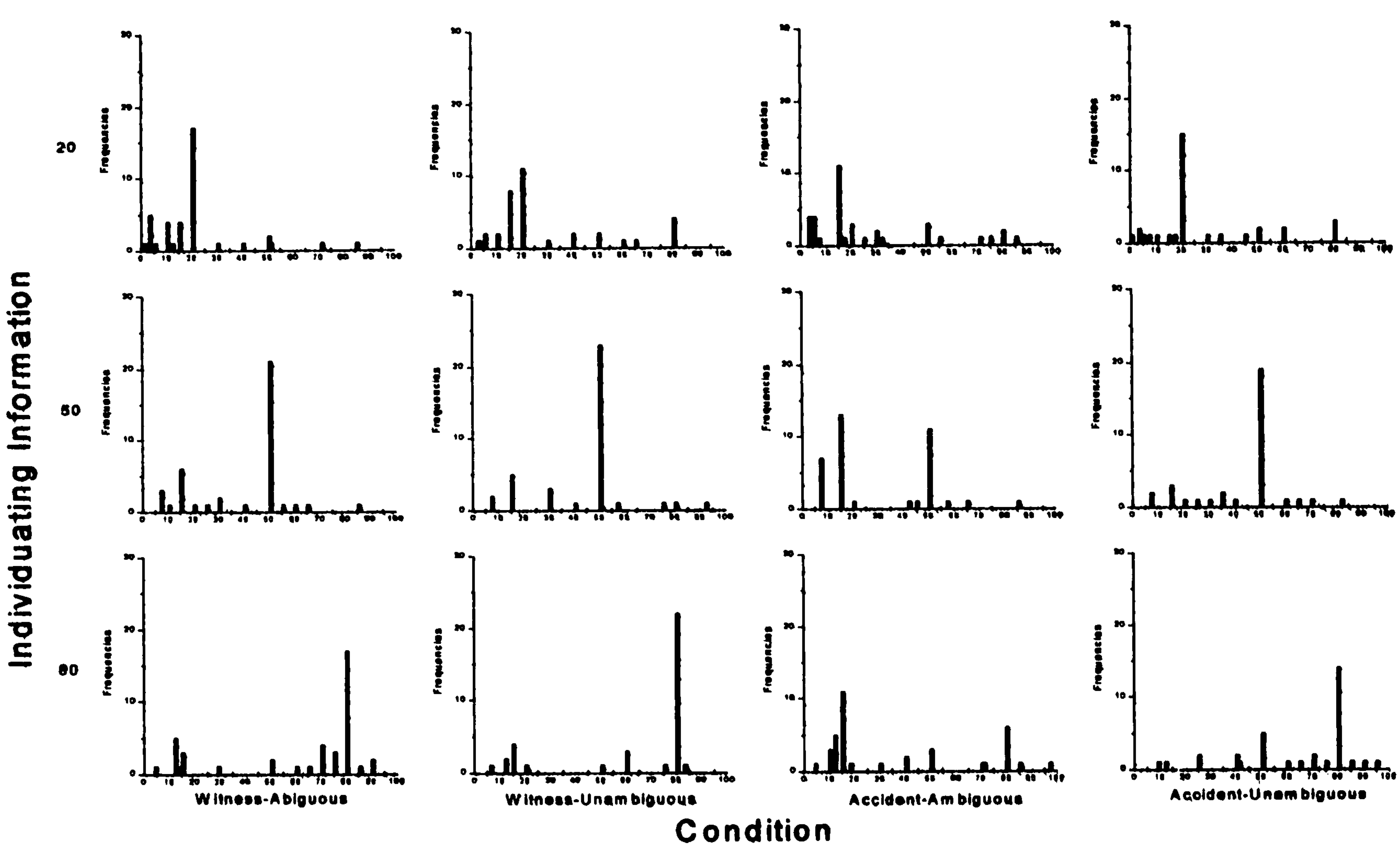

Figure 13: Histogram of raw probability judgements for the Low Base Rate (15) conditions of Experiment 3.

At first glance, the finding that Individuating information had less impact when the probability to be assessed was conditional on the accidental status of the cab embedded in an ambiguous formulation seems to contradict the conversational analysis that was proposed based on the maxim of manner. Certainly, since the ambiguous formulation contains the potential for confusion between the computation of a / (a +c) and that of a / (a + b), this condition was not expected to elicit more normative answers. At the same time however, according to the maxim of quality, the accidental status of the cab was expected to less readily elicit attributes of sincerity, honesty, reliability, and competence. This could have caused the participants in this condition to be less attracted to the information contained in the percent correct colour predicted by the accidental status of the cab. They had to begin searching for other solutions. There is clear evidence in Figure 12 and Figure 13 that participants in this condition were trying other ways of solving the problem. These figures show the distribution of the raw probability judgements. Working from left to right are the four between-subject conditions P(B|W) Ambiguous, P(B|W) Unambiguous, P(B|A) Ambiguous, and P(B|A) Unambiguous. Going down the panels from top to bottom are the increases in Individuating Information (Low = 20, Medium = 50, and High = 80). Comparing the third column with the other three one can clearly notice a difference in probability distributions. In the P(B|A) Ambiguous condition, the distribution is multimodal. What the modes correspond to is speculative yet the most frequent responses seem to approach the Base-Rate information (15 in the Low Base-Rate problems and 85 in the High Base-Rate problems). Another set of responses seems to match the Individuating Information; and a final set of responses might correspond to the conjunction of the base-rate information multiplied by the

the conjunction of the base-rate information multiplied by the individuating information. For example, in the LBHI problem (base rate of 20 and individuating information of 80), the conjunction strategy results in a judgement of 12. Nevertheless, whatever the strategies were, there were many of them, unlike the more unimodal pattern seen in the other conditions.

Why was the multimodal pattern not obtained in the P(B|A) Unambiguous condition as well? Maybe because the disambiguation did not work as intended. It is possible that the so-called unambiguous formulation simply reinforced the notion that the percent correct colour predicted by the accidental status of the cab was a valuable piece of information (so they were attracted by it in the same manner as to the witness testimony) and participants still mistook the computational ratio a / (a + c) for a / (a + b). Overall, the results of Experiment 3 shows that a combination of the context (witness vs. accidental status) and wording (ambiguity) of the problem induce differential judgement strategies - strategies that are a function of the degrees of uncertainty in the problem and the meaning of words in conveying the uncertainty/reliability of the information being presented. This remains consistent with a conversational analysis of the problem structure (e.g., Grice, 1975; Hilton, 1995b; Macchi, 1995; Schwarz, Strack, Hilton, & Naderer, 1991), highlighting the importance of how information is conveyed and how it is open to misinterpretation. Accordingly, a truly unambiguous problem formulation is required for the proper (normative) encoding of probability information. To demonstrate this is the goal of the next and last experiment.

Experiment 4

In Experiment 4, a new problem format was designed to present the necessary numerical information in a clear and conceptually sound manner. To achieve this, the numerical information contained in a word problem was always presented in a two-by-two table format similar to Table 1.

The experiment was also designed to test the hypothesis that once a sound conceptual structure was achieved and matched the architecture of the human cognitive system, then the amount of information presented in a problem would add to the degree of its complexity, thereby influencing the strategies used to solve the problem (Kleiter, Krebs, Doherty, Garavan, Chadwick, & Brake, 1997) and the likelihood of a correct answer. This would account for the discrepancies between actual judgements and normative values. The experiment was a mixed factorial design crossing two between-subject dimensions (two levels of Witness/Accidental status and four levels problem difficulty) with two within-subjects dimensions (two levels of base rate and three levels of Individuating information). The four levels of problem difficulty were manipulated within-subject by providing either percentages or frequencies as data in the cells of the two-by-two table. In one condition (All % and Base), the information in all four cells was provided as percentages along with the two base rate frequencies (number of Blue and Green cabs). The base rate frequencies were provided to allow people to solve the problem via the a / (a + b) calculations, without recourse to Bayes' theorem. In this condition, arriving at a normative judgement first entails calculating the frequencies in cells a and b by multiplying the given percent by the base rate. For instance, if the proportion correct of the witness testimony or accidental status is given as 80% and there are 15 Blue cabs, multiplying 80% by 15 yields a frequency of 12 Blue cabs that were reported to be blue (or 12 Blue cabs that were involved in accidents). Cell b is obtained by multiplying 20% by 85 resulting in a frequency of 17 Green cabs that were reported to be blue (or involved in accidents). Once these two frequencies have been determined, the normative solution for all problems is cells a / (a + b). In a second condition (All Freqs), the four cells, and only the cells, were provided as frequencies. Although not required to solve the problem, the base rate frequencies could be calculated by the participants by summing cells a + c and cells b + d respectively. In a third condition (2 Freqs & Base), only cells a and b were provided as frequencies along with the base rates. Here again the base rate frequencies are not required but were provided in order to present the same number of candidate ratios to choose from as in condition All % & Base. Finally, the fourth condition (2 Freqs) presented only cells a and b as frequencies. Thus the normative solution to all the problems presented in Experiment 4 does not require understanding nor calculation of Bayes' theorem and can be arrived at by computing the ratio of cells a / (a + b) and optionally, converting this proportion to a percentage by multiplying by 100. Accordingly, the problem is most difficult to solve in condition All % & Base because participants must first select the relevant pieces of information among a larger data set with a risk of making one or more wrong choices. Then the information must be converted from percentages to frequencies and, finally, the ratio a / (a + b) must be calculated. Condition All Freqs is intermediate in that it eliminates the need to convert percentages but maintains the risk of selecting inappropriate data to enter into the final ratio. Condition 2 Freqs & Base Rates is less difficult because it reduces the risk of selecting wrong data and it eliminates the conversion of percentages to frequencies while maintaining a large choice of possible ratios to compute [e.g. a / (a + b), a / base, a / (a + c), etc.]. Condition 2 Freqs is the easiest because it provides the fewest pieces of information to choose from, requires no conversion from percentages, and only requires the calculation of the ratio a / (a + b). Note that although this condition is the easiest, it is not completely free of potential mistakes. For instance, participants might calculate b / (a + b), make errors in the arithmetic, or just fail to understand that a ratio must be calculated.

If what is really determining a normative use of probabilistic information is the match between the human cognitive architecture and the format of the problem, then judgements made in the frequency formats should elicit probability judgements that more closely approximate normative expectations, more so than in the problem formats of the three previous experiments where the information was presented in percentages (single-event probabilities). In addition, if problem difficulty expressed in number of algorithmic steps required to solve the problem is a central determinant of performance, then performance accuracy should rank order according to the four conditions of complexity.

If judgements are closer to normative expectations in the Accidental Status condition and less normative in the Witness conditions, these results would support the Conversational Conventions hypothesis based on the maxim of quality. Alternatively, if the problem structure is the crucial determinant of solution strategies then we would expect no difference between the Witness and Accidental Status conditions.

Method

Participants. One hundred eighty-one undergraduate students enrolled in first year Psychology courses at the University of Ottawa served as participants in this experiment.

Materials. In this experiment, the information is always presented in a table format. Table 5 lists examples of all the tables that were presented to the participants. The tables are grouped in the figure for ease of presentation but were always presented singly during the experiment.

Design and Procedure. As with Experiment 3, two levels of Base Rates (15 and 85) and three levels of Individuating Information (20, 50, and 80) were presented within-subjects. One of the two between-subjects factors involved 2 levels as the conditional source of information (Witness versus Accidental Status), while the other involved 4-levels of problem difficulty (the information presented as: All %, All Frequencies, 2 Frequencies & Base Rates, or only 2 Frequencies). This resulted in 2 x 3 x 2 x 4 mixed factorial design with 8 between-subject conditions and 6 within-subjects conditions.

As in the three previous experiments, the colour of the cab in the question was always counterbalanced with three problems asking about a Blue cab and the three remaining questions asking about a Green cab. The colour of cab asked in the question was also counterbalanced for the high and low Base Rate conditions. For instance, in the All %-Witness condition 12 problems were generated in all: 3 low base rate Blue cab questions, 3 low base rate Green cab questions, 3 high base rate Blue cab questions, and 3 high base rate Green cab questions. Thus, a problem booklet consisted of 6 problems placed in random order 3 low base rate and 3 high base rate problems (counterbalanced for question of cab colour and base rates) - one problem from each of the 6 within subject problems types.

Table 5: Experimental conditions for Experiment 4.

All % - Witness

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

		True Cab Colour: Blue	True Cab Colour: Green
Witness Testimony	blue	80%	20%
	green	20%	80%
		15	85

What is the probability that last night's cab was Blue given that a witness identified the cab as blue?

All Frequencies - Witness

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

		True Cab Colour: Blue	True Cab Colour: Green
Witness Testimony	blue	12	17
	green	3	68

What is the probability that last night's cab was Blue given that a witness identified the cab as blue?

2 Frequencies & Base Rates - Witness

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

		True Cab Colour: Blue	True Cab Colour: Green
Witness Testimony	blue	12	17
		15	85

What is the probability that last night's cab was Blue given that a witness identified the cab as blue?

2 Frequencies - Witness

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

		True Cab Colour: Blue	True Cab Colour: Green
Witness Testimony	blue	12	17

What is the probability that last night's cab was Blue given that a witness identified the cab as blue?

All % - Accidental Status

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

Accidental Status		True Cab Colour: Blue	True Cab Colour: Green
	Accident	80%	20%
	Non-Accident	20%	80%
		15	85

What is the probability that last night's cab was Blue given that it was involved in an accident?

All Frequencies – Accidental Status

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

Accidental Status		True Cab Colour: Blue	True Cab Colour: Green
	Accident	12	17
	Non-Accident	3	68

What is the probability that last night's cab was Blue given that it was involved in an accident?

2 Frequencies & Base Rates – Accidental Status

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

Accidental Status		True Cab Colour: Blue	True Cab Colour: Green
	Accident	12	17
		15	85

What is the probability that last night's cab was Blue given that it was involved in an accident?

2 Frequencies – Accidental Status

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

Accidental Status		True Cab Colour: Blue	True Cab Colour: Green
	Accident	12	17

What is the probability that last night's cab was Blue given that it was involved in an accident.

Results and Discussion

Mean probability judgements are shown in Figure 12. The figure reveals that mean probability judgements increased across panels, corresponding to a significant main effect of Individuating Information, F (2, 346) = 134.65, MSE = 382.03, η^2 = .438. Mean judgements also increased within panels, corresponding to a significant main effect for Base Rates, F (1, 173) = 237.49, MSE = 821.19, η^2 = .579. These two effects replicate Experiments 1 to 3.

Comparing the top three panels with the bottom three, it is clear that there was no difference between conditional probability judgements depending on a witness statement or those depending on the accidental status of the cab. In addition to the lack of main effect for the Witness/Accidental status factor, no significant higher order interactions involving Witness/Accidental Status condition were found, thus failing to support any differences in probability judgements that may be attributable to a Conversational Analysis in terms of the maxim of quality.

The most notable pattern of results in Experiment 4 concerns the 4 levels of problem difficulty. The pattern of judgements made in the All % conditions is noticeably influenced by changes in Individuating Information and rather little by the changes in Base Rates. By comparison, judgements made in the three frequency conditions appear more influenced by base rates (steeper slope than All % & Base within each panel) and less influenced by individuating information (lesser average change than All % & Base from panel to panel). This observation is statistically supported by two simultaneous interactions, one involving Difficulty and Base Rate, F (3, 173) = 23.173, MSE = 821.17, η^2 = .287, the other involving Difficulty and Individuating information, F (6, 346) = 6.21, MSE = 382.033, η^2 = .097. To better visualise these interactions, they are plotted separately in Figure 13 and Figure 14.

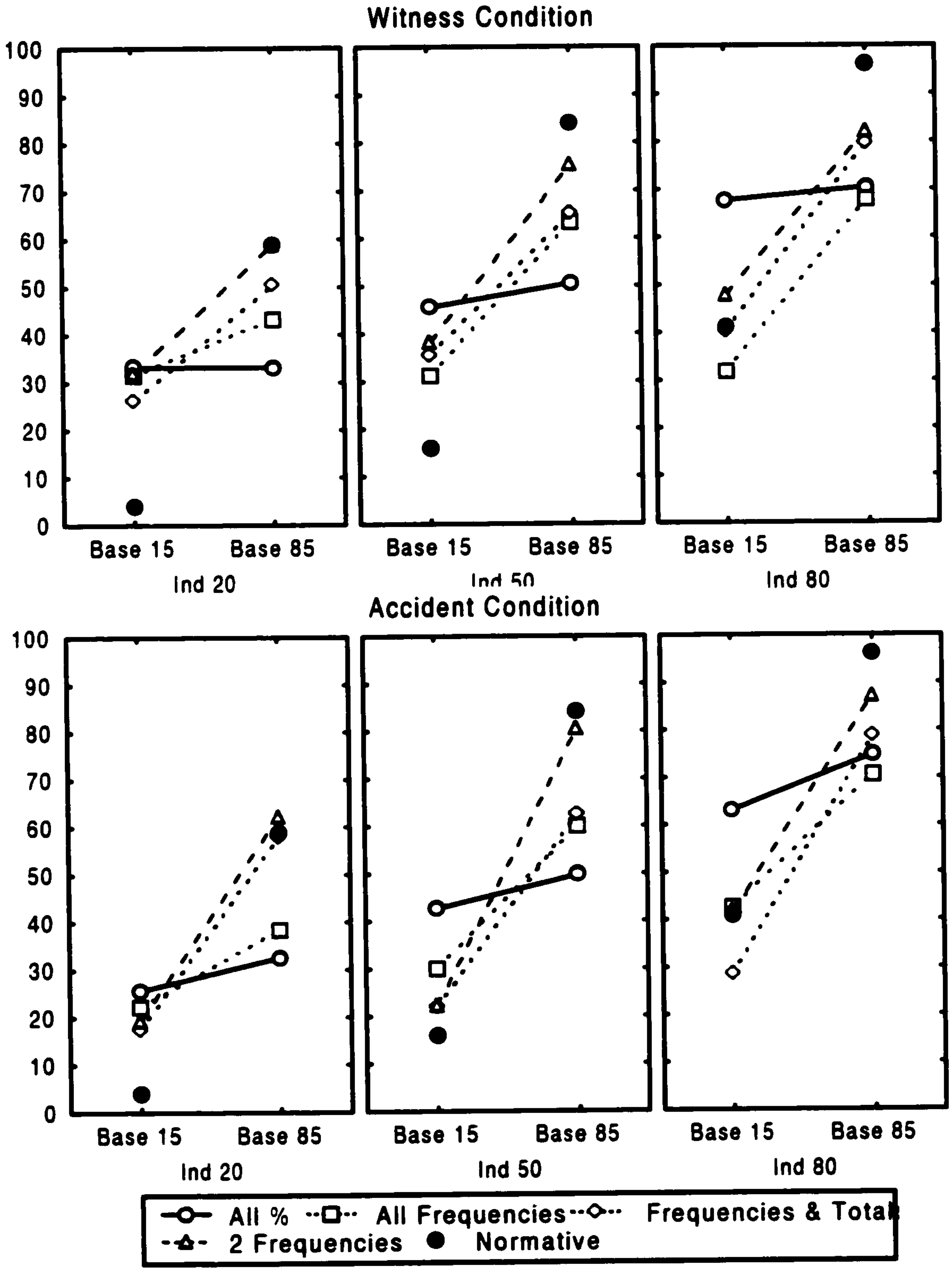

Figure 12: Results of Experiment 4

In Figure 13, level of difficulty is plotted against base rate. The better the discrimination between base rate 15 and base rate 85, the larger the distance between the judgements. Thus the sensitivity to base rate information is reflected in the steepness of the slope across base rates. The figure shows not only that the All % & Base condition is less sensitive to the base rate but also that the three frequency conditions vary in that sensitivity in a manner that corresponds to the level of difficulty. The slope is flattest for the most difficult problem (All % & Base) and becomes steeper across conditions (All Freqs and 2 Freqs & Base) to reach its steepest in the easiest problem (2 Freqs).

Individuating Information also influenced mean probability judgements in a manner that was complementary to the base rates. Figure 14 shows that judgements increased as a function of

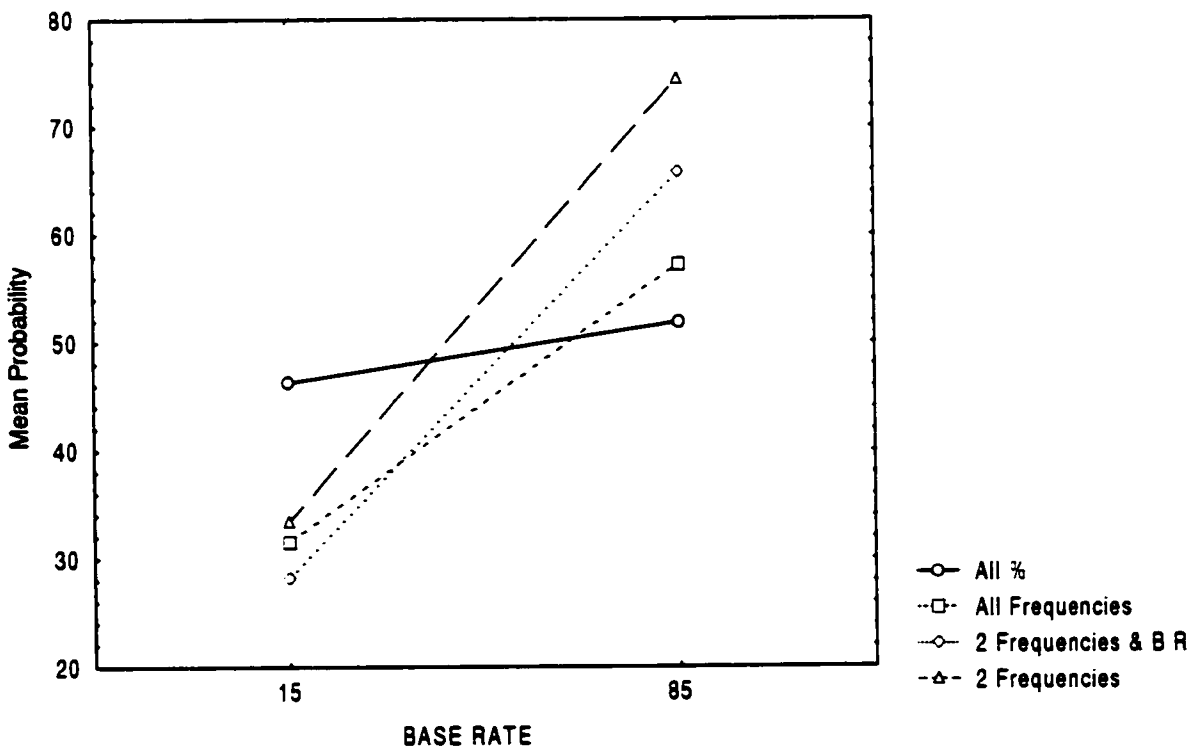

Figure 13: DIFFICULTY by BASE RATE interaction.

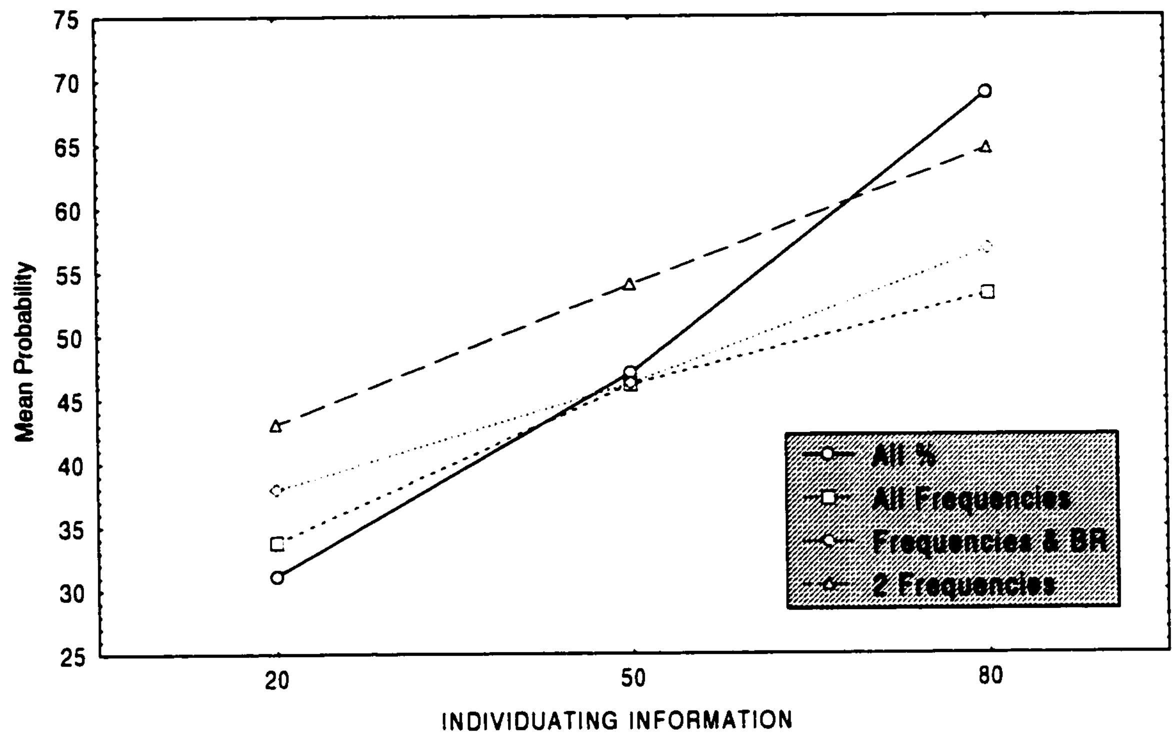

Figure 14: DIFFICULTY by INDIVIDUATING information interaction

the change in Individuating Information. However, judgements were differentially influenced according to the difficulty of the problem. As summarised more clearly in Figure 16, judgements made in the All % conditions were more strongly influenced by the Individuating Information since the slope of changes in this dimension is steeper for the All % conditions than all other conditions. The same slope becomes flatter and flatter as the problem difficulty decreases

The raw probability judgements depicted in Figure 15 to Figure 18 show the individual judgement pattern underlying the ANOVA results. The judgements in the All % condition for both the Witness and Accidental Status in the Low and High Base Rate conditions matches very closely the pattern of responses in obtained in Experiments 1, 2, and 3. In the top three panels of Figure 19 to 22, the modal response corresponds to the Individuating Information, either 20, 50, or 80. In the frequency conditions, a different pattern emerges, one where the modal response no longer solely corresponds to the Individuating Information. For instance, in the 2 Freq conditions, the modal respond is the normative response (a / (a + b)). This pattern holds true for the Witness and the Accidental Status conditions as well as the Low and High Base Rate conditions. Further, the amount of dispersion is less in this condition than in any of the other conditions. As the task is made more complex by introducing more steps in the cognitive process, judgements deviate further from normative expectations and acquire more variability. This is true irrespective of the source of the conditional (Witness or Accidental Status).

The results of this experiment are consistent with those in the previous three experiments. Problems posed in the percentage format (which includes most if not all the problems posed by Tversky and Kahneman, 1974) result in judgements that are primarily influenced by Individuating Information and only secondarily influenced by the Base Rate information. However, changing the problem format to frequencies conveyed in a contingency table, individual judgements become more normative. In the three frequency conditions, judgements were most sensitive to the changes in Base Rates (η^2 = .579) and secondarily to changes in Individuating information (η^2 = .438).

Note that in this experiment, the expression "sensitivity to base rates" is used descriptively to indicate changes in participants' judgements matching changes in problem data as created by the experimenter. In actuality, participants may or may not compute base rates per se. Although it is feasible for participants to compute base rates everywhere except in the 2 Frequencies condition, we do not know for sure whether they do compute them. In addition, the fact that the alternative algorithm of computing a / (a + b) is available everywhere, and bypasses the need to compute a base rate, suggests that this is what participants might do in general. On the other hand, computing base rates when that is feasible may contribute to the complexity of the problem in that participants have to determine whether that computation is necessary or even useful.

Taken together, the two complementary interactions of Difficulty by Base Rate and Difficulty by Individuating Information indicate that the amount of deviation from normative expectations is a clear function of problem difficulty as indexed by the number of computational steps involved in the solution. This finding clarifies an ongoing issue in the existing scientific literature on the use of base rate information, namely what controls the extent to which base rate information is taken into consideration in solving a given problem. This finding also goes a long way towards detailing the exact nature of the cognitive process(es) subsumed under the label "heuristics" by Tversky and Kahneman. To the extent that this experiment has identified steps in a rational process of problem solving, it points towards a process continuum, where mistakes can be made, and replaces the poor explanation that heuristics have been. In the heuristic view, the discovery of a new judgement bias often leads to the post-hoc offering of a new heuristic to account for the bias, in effect merely pushing the explanation one level back. In the problem complexity view, less and less normative answers are predicted according to the number of stages in the problem resolution process where mistakes can be made. This is a more integrated view based on a steady system of cognitive processes that can be defined a priori.

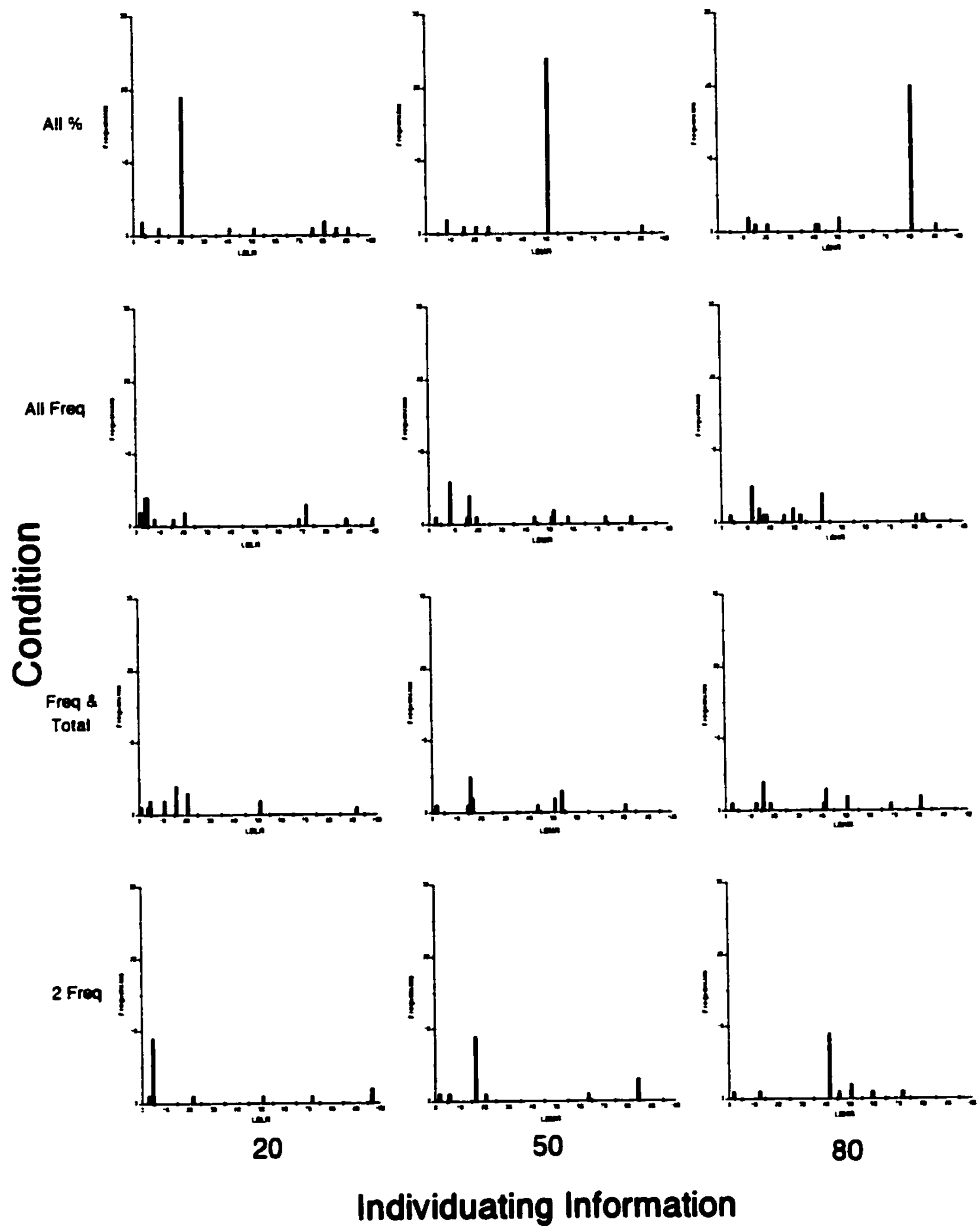

Figure 15: Histogram of the raw probability judgements made in the Low Base Rate – Witness condition in Experiment 4.

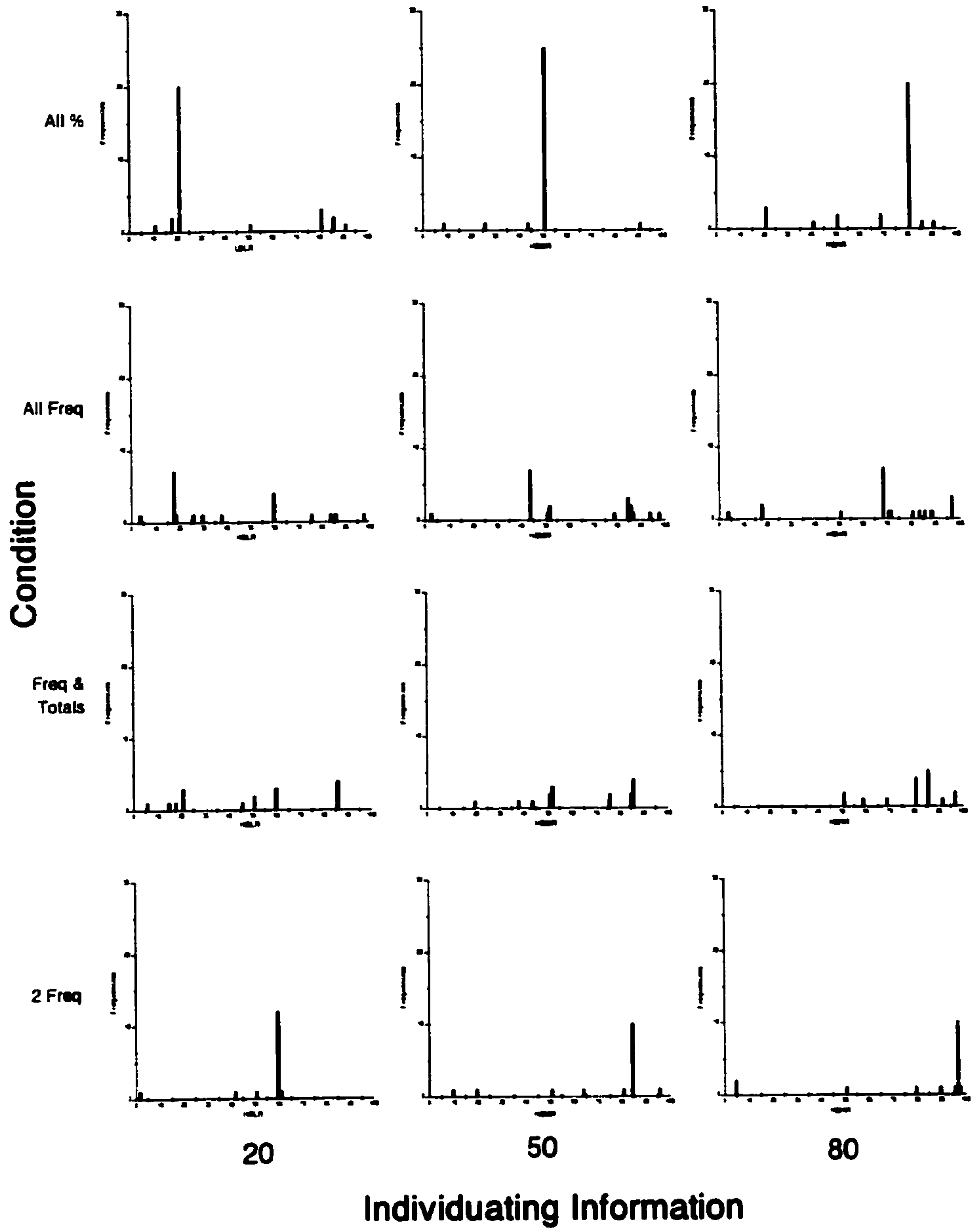

Figure 16: Histogram of the raw probability judgements made in the High Base Rate – Witness condition in Experiment 4

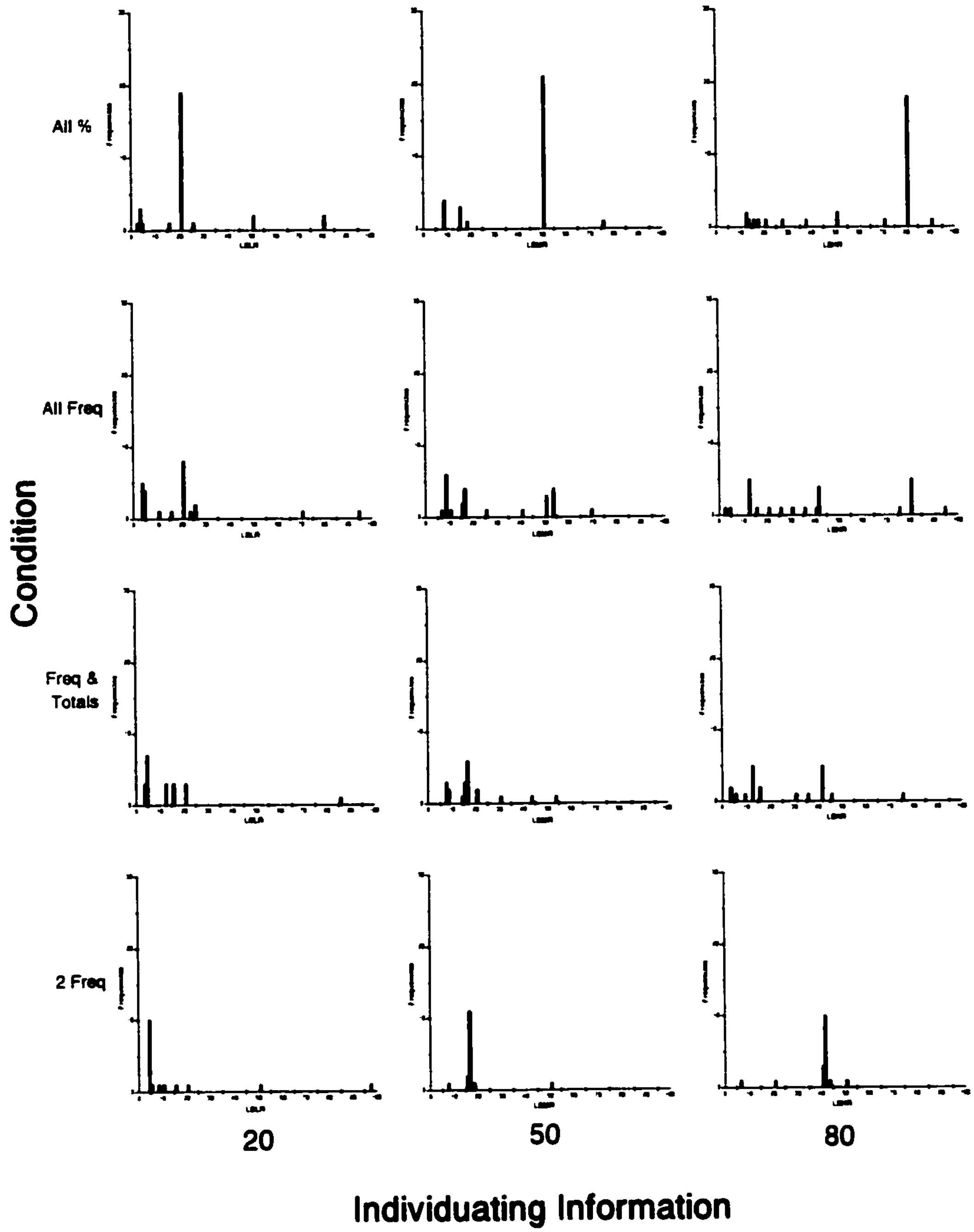

Figure 17: Histogram of the raw probability judgements made in the Low Base Rate – Accident condition in Experiment 4.

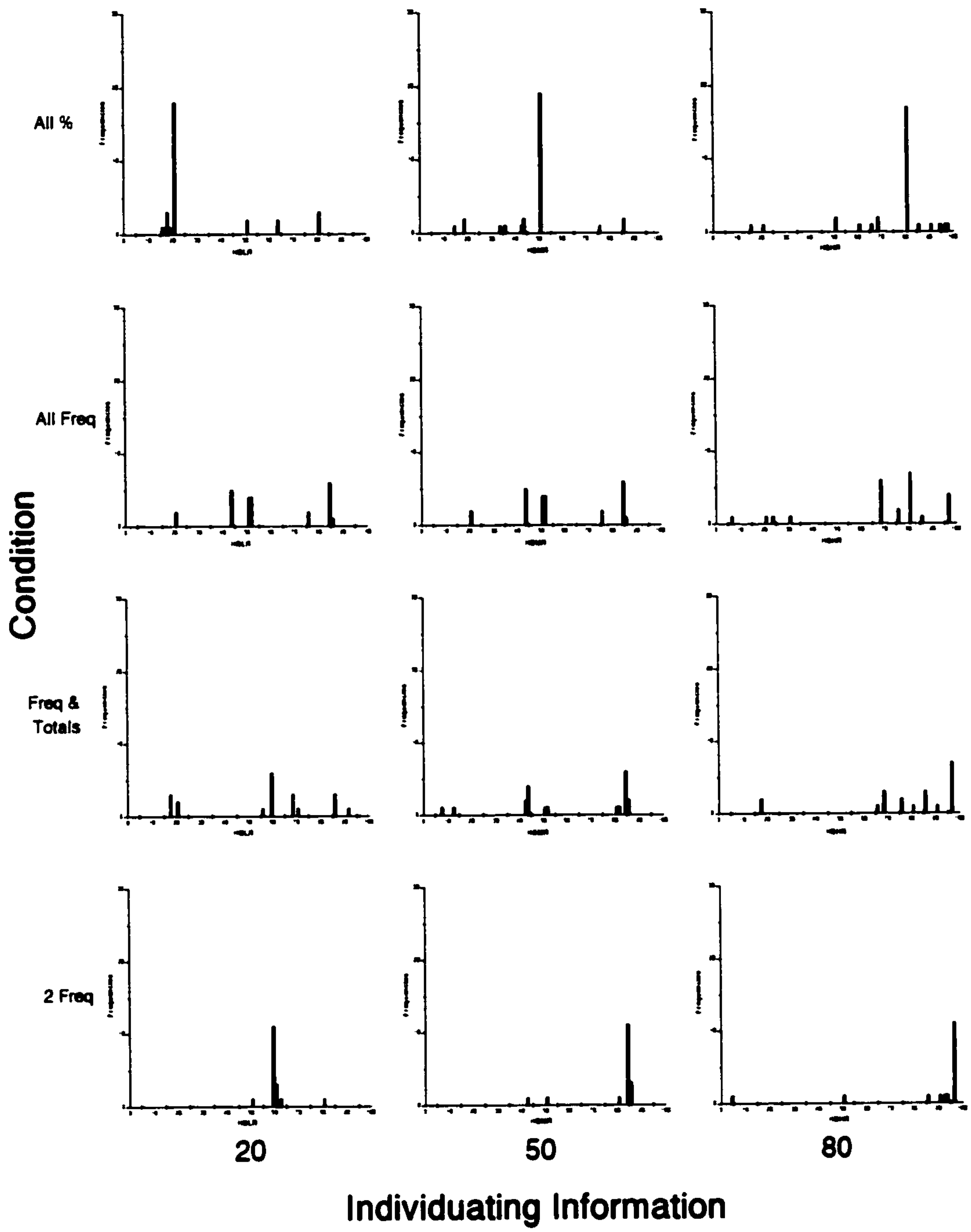

Figure 18: Histogram of the raw probability judgements made in the High Base Rate – Accident condition in Experiment 4.

General Discussion

The four experiments reported in this thesis extend our understanding of how problem-solvers use numerical information in probabilistic reasoning tasks, with particular emphasis on how they use and misuse base rate information. In each experiment, a series of problems varied the levels of base rate and individuating information across experimental conditions. All experiments used variations of the cab problem as described by Tversky and Kahneman (1982c). The variations in format of presentation included: on-line (trial-by-trial), off-line (word problem), percentages, and frequencies.

Empirical Observations concerning Base Rate neglect

The well known results of Tversky and Kahneman (1974; 1980; 1982c) and the many subsequent studies that have been inspired by their research program initially led researchers to make strong statements about how people neglect base rate information in probability problems (Bar-Hillel, 1980; 1983; 1990). But other research results were mixed on the extent to which base rates are actually neglected (Koehler, 1996). In this thesis, all four experiments clearly showed that participants were reliably and sizeably sensitive to base rate information. Though individuating information remained the most salient, variations in base rates definitely impacted on judgements.

At an empirical level, this thesis has shown that the on-line (i.e., through direct experience trial-by-trial) versus off-line (word problem) presentation format of probabilistic information is not a crucial determinant of base rate sensitivity since this factor never interacted reliably with base rate level. There is a faint possibility that the on-line format had the same end effect as making the base rates causally relevant (Bar-Hillel, 1980) to the experimental task. Recall that when witness reliability is 80 and base rate is 15, the normative response is 41 and that the studies manipulating the relevance of the base rate information (e.g., Bar-Hillel, 1980; 1983; 1990; Tversky & Kahneman, 1980; 1982c) found the modal response to be 60. This is the same level of response obtained in the two on-line experiments here (Experiments 1 and 2). Experiment 1 also obtained this level of response off-line, possibly because of a within-subject carry-over from the on-line condition. But this observation is weak compared to the true within experimental comparisons.

Though the results of the four experiments show an effect of both pieces of information on judgements, responses were still not normative. This raises questions as to how the various pieces of information provided in any problem statement are integrated together if at all. To begin to answer this, we must turn to the prevalent theoretical models and examine what they have to offer. Three approaches are considered in detail: 1- the heuristics approach (Kahneman et al., 1982), 2- the frequentist approach (Cosmides & Tooby, 1996; Gigerenzer, 1998), and 3- the conversational analysis approach (Grice, 1975; Hilton, 1995b). The results of the first three experiments address the theoretical predictions made by each approach and suggest that a fourth theoretical viewpoint based on task complexity, or the number of cognitive steps involved in arriving at a solution, should supersede them (Kleiter et al., 1997).

Heuristics

The heuristics approach suggests that when base rates are shown to impact judgement it is because they are integrated with the individuating information in an "adjustment" mechanism. In this view, the numerical value of the individuating information is first selected as an anchor (a selection process which requires a separate heuristic to be accounted for) which is then modified according to the numerical value of the base rate. Based on the numerical value of the final judgements, it is tempting to infer that the nature of the adjustment was multiplicative (Jeffereys, 1961). Indeed, the examination of the individual response patterns in the P(B|A) Ambiguous condition of Experiment 3 is consistent with this. But the inference is post-hoc and unable to account for the more general findings. For instance, the multiplicative pattern was not observed in all conditions in all experiments although base rate sensitivity was observed everywhere. What was the integration formula in the other conditions? In addition, it is well known from the research investigating the "conjunction fallacy" that people often fail to see that the probability of joint independent events is multiplicative. Rather they seem to think that they are additive. Why then should people use a multiplicative information integration algorithm in conditional probability problems but not in conjunction problems? The heuristics approach might propose that participants have a wide variety of heuristics at their disposal and rely on different ones at various times. But that falls very short of enabling us to predict what people will do when.

Frequentist Format

As an alternative to heuristics, some researchers (Cosmides & Tooby, 1996; Gigerenzer, 1991b; 1996) have proposed that the mind has evolved according to a natural selection process whereby it has come to rely on an algorithm suited to the use of naturally occurring frequencies. These authors postulate that the human cognitive system has adapted to naturally process information in a frequentist format and have presented data showing that judgements are closer to normative expectations when information is presented in frequencies (Gigerenzer & Hoffrage, 1995).

According to this view, base rate information should be used appropriately as longs as it is presented as a frequency. However, this was not the case in the on-line conditions presented here. Clearly there were no differences in judgement patterns between the on-line and off-line conditions of Experiment 1. The presentation of frequencies via an ecologically valid, trial-by-trial, direct experience format was insufficient to produce more normative judgements than probability-based word problems.

Frequentist authors have not attempted to determine how base rates are used naturally. Rather, they changed the wording of their question and directly asked participants to provide an answer in the form "X cabs out of N". This method imposes a specific algorithm on the participants which might conceal another algorithm that would have been used had the people been freer to do so. So we do not know if the ratio X out of N is natural.

A third and final shortcoming of the frequentist approach is that it does not explain why answers are still incompletely normative even when problems are presented as frequencies and the need to search for a ratio of two target frequencies is explicitly stated for the participants. In Experiment 4, it was shown that increased normativeness spans a continuum of frequency and probability data and is not restricted to one type of information. In that experiment, the short frequency version problems produced judgements closest to the norm. This portion of the finding is consistent with the frequentist approach. However, the frequentist approach does not offer a comprehensive explanation for the graded rate of success in solving the various problem formats in Experiment 4.

Conversational Discourse in Probabilistic Reasoning

The principles of conversational analysis can be applied to performance in probability problems to determine which element of information, base rate or individuating information, is expected to have the greater weight. Beyond this, the information integration is not explicitly specified but is clearly assumed to be rational, hence normative.

Tests of conversational analysis principles were more prominent in Experiments 2 and 3. Specifically, the maxim of quality, which pertains to the truth value of the informational source, was tested by comparing two conditional sources of individuating information: the witness' testimony and the cab's accidental status. According to the maxim of quality, the truth value of any witness statement would be an issue to contend with hence drawing attention to that piece of information and driving answers towards it. Accordingly, the truth value of the source of information really should have been a consideration only in the witness testimony condition as it explicitly brings up the reliability of the witness. In the accidental status condition, the non-human nature of the source makes it more on par with the base rate and should not so strongly attract attention. In the on-line presentation format of Experiment 2, no significant differences were found between the two conditional sources thus failing to support the importance of the maxim of quality. Since the three conditional sources were run between-subjects there was no possibility of cross condition carry over in that experiment.

In Experiment 3, the maxim of manner was explicitly tested by creating so-called ambiguous and non-ambiguous problem statements. According to the maxim of manner, a problem statement should be as unambiguous as possible and, consistent with this, pieces of information emphasised in the problem statement would be assumed to be more important. In the Tversky and Kahneman problem formulation, the individuating information is emphasised by quoting its reliability level whereas the base rate is simply provided as is. This formulation is termed ambiguous because it has the potential of implicitly inducing the problem solver to decide that the individuating information is more important. A less ambiguous formulation in which both the individuating information and the base rate are given equal prominence and, in addition, in which the nature of the numerical value is as clear as possible, should yield more normative answers because the problem solver would not be as biased by the manner of the formulation. Indeed ambiguity did have a significant effect on judgements. However, contrary to the foregoing conversational analysis, judgements were impacted the most by base rate in the Ambiguous-

Accidental status condition and not in the unambiguous one. This finding contradicts the conversational analysis based on the maxim of manner.

At the same time however, according to the maxim of quality, the accidental status of the cab was expected to less readily elicit attributes of sincerity, honesty, reliability, and competence. This could have caused the participants in this condition to be less attracted to the information contained in the percent correct colour predicted by the accidental status of the cab and make them begin to search for other solutions. So a conversational analysis cannot be rejected outright. It does not provide an account for the graded normativeness of the judgement pattern of Experiment 4, and its predictions based on the interplay of the maxims of quality and manner are not unequivocal. Yet it does emphasise the overriding importance of problem formulation which was shown to matter most in the current experiments.

In addition to problem formulation per se, other researchers have made the point that how participants interpret the information given is not always how it was meant to be interpreted. This is conversational analysis under another name. For instance, some research participants were found to confuse the posterior probability P(H|D) with the inverse conditional P(D|H) (e.g., Braine, Connell, Freitag, & O'Brien, 1990; Koehler, 1996). In the cab problem this amounts to interpreting the witness reliability information as the posterior probability (answer). Thus the posterior probability sought in the question "What is the probability that the cab was Blue given that a witness reported it being blue [P(B|b)]," is confused with the witness reliability information [P(b|B)] provided. If participants confuse these two pieces, then it is no wonder that the modal response rate is 80 - corresponding to the individuating/indicant information.

The results of Experiments 3 and 4 are consistent with the general notion of the importance of problem context (Grice, 1975). Attempting to understand judgement biases in probabilistic inference, Machi (1995) analysed the verbal structure of problem questions and concluded that the use of base rates is linked to the effective transmission of the relationship between the pieces of information conveyed in the problem. In the cab problem, this corresponds to the relationship between base rates and the indicant (witness/accidental status) information. Machi argues that the neglect of base rates is due to the ambiguity of the information that is produced by the structure of the problem, ambiguity due to the absence of a partitive structure. This amounts to a confusion between the two conditional probabilities, P(B|b) and P(b|B) which in turn amounts to the same thing as asserting the consequence (Braine et al., 1990). Thus, if participants see the specific information as the required posterior probability, then there is no need for them to attend to base rates. If this is the case then it should be possible to improve judgmental responses by removing the ambiguity in the text by referring to the percentages of the witness' identification to the base rate (Macchi, 1995). To test this, she replaced the witness information with the following: "A witness recognised as Blue 80% of the Blue cabs and mistook 20% of Green cabs for Blue ones. Among those cabs indicated as Blue what is the percentages of cabs that really are Blue?" (p. 207). Results of probability judgements of this formulation were significantly different from the classical cab version since between 70% and 74% of the judgement responses considered both the individuating information and the base rates. Thus, Macchi concludes that when the percentages in the specific information refer to the correct referent (e.g., the percentage of Blue cabs reported as being blue), then there is no confusion between the two conditional probabilities (i.e., between the individuating information and the posterior probability). Clearly stating the individuating information in this explicit manner conveys the relationship of independence between the base rates and the individuating information (Macchi, 1995). She goes on with her analysis by suggesting that the partitive element transmits the idea that the individuating information is related to a particular property (in the cab problem this is the property of being recognised as being blue), defining a subset that is not to be considered separately, but as a global set in which the base rate refers (Macchi, 1995, p. 204).

In light of these results Macchi (1995) relates the frequentist approach espoused by Gigerenzer with the Bayesian approach in terms of how the information is transmitted. Rather than the mind being frequentist in its representation, Macchi argues that what seems to occur is that the frequentist task clarifies the relationship between the different pieces of information. That the relative frequency of an event defines its probability in a non-ambiguous manner, that is, it clearly expresses the independence of the base rate and individuating information. For example, in the frequency format of the cab problem, the statement "in a sample of 100 cabs, 12 out of every 15 Blue cabs are identified as blue cabs" clearly indicates the independent relationship between the base rates (that 15% of the cabs in the city are Blue) and the individuating information (that 80% of the Blue cabs are correctly identified as blue) and at the same time specifies the necessary information needed to arrive at a normatively correct solution.

However, even with ambiguity controlled for as in Experiment 3 where the unambiguous condition respected the partitive structure in the problem statement, less than normative answers were obtained. Conversely, an ambiguous statement did bring participants to pay more attention to base rates. The difficulty lies in trying to analyse problem solving at such a high level of processing and ignoring or subsuming under the global design all the finer grained mental operations that are required to process the information. Therefore, a fourth theoretical approach is presented below which takes this into consideration.

Task Complexity and Cognitive Demands

Thus far, we have examined the role of heuristics, frequency formats, and conversational conventions in judgements under uncertainty. Up to this point, no single theory can account for all the results in Experiment 1, 2 and 3. Experiment 4 was designed to test the hypothesis that what really is at work is the complexity of the judgement task. The more steps involved in arriving at a solution, the greater the likelihood for computational errors and the more deviation from normative expectations. The results of Experiment 4 show exactly this. As the task structure becomes more complex, from presenting only two required pieces of information in the 2 Freq conditions to the All % conditions, responses became less normative. This gradation was obtained simultaneously with a replication of the main effects of individuating information and base rate obtained in the first three experiments. Further, this held true irrespective of whether the conditional source was the witness or the accidental status.

As the cognitive complexity of the task becomes simpler, judgements become more normative. In the problems presented in Experiment 4, all the information is presented in a table format. In this case, participants do not require the formal knowledge and understanding of Bayes' Theorem but instead can arrive at a normative solution in the form of cells a / (a + b). What is required in the All % condition is that the raw probabilities be converted to frequencies prior to implementing the solution strategy. In the three frequency conditions, this conversion has already been done, but participants must select the appropriate information to use and then arrange it in the form of a / (a + b). In the simplest case (2 Freq condition), only these two cells are provided, thus participants only need to put the information in the proper ratio. The most frequent response in the 2 Freq condition was the normatively correct judgement. As more information is provided, as in the other two frequency conditions, the task requires participants to determine what the necessary information is and apply it to the algorithm a / (a + b). Participants still made mistakes occasionally even in this simplest condition. But if a problem is defined very narrowly, there are few possibilities for mistakes. In the 2 Freq condition, it was still possible to calculate a wrong ratio (e.g. b / (a + b)) or to get the arithmetic wrong. In the more complex problems, there were more ratios to choose from and more calculations to be done so that, on average, answers became less normative but for a variety of mistakes rather than strictly because of base rate neglect.

Summary

The four experiments presented in this thesis have tested three distinct theoretical frameworks in explaining the use and neglect of base rate information in probabilistic reasoning problems. This was done in on-line and off-line problem formats. None of manipulations made in Experiments 1, 2, or 3 were sufficient in producing a normative pattern of responses. It was not until problem complexity and the number mental operations were taken into consideration that a pattern of normative judgements emerged. Thus, perhaps what is at the "heart" of all probabilistic reasoning is the number of mental operations required to arrive at a normative judgement. Tasks that present information in a probability format place greater cognitive demands and require more work than tasks using a frequency format.

If there is one lesson to be learned from this research effort, it is that cognitive scientists should be very careful in making inferences about cognitive processes based on very macroscopic experimental manipulations unless they can guarantee that the microscopic processes are kept constant… a danger that has been and clearly remains with all of behavioural science.

References

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. Journal of Personality and Social Psychology, 35(5), 303-314.

Allan, L. (1980). A note on measurement of contingency between two binary variables in judgment tasks. Bulletin of the Psychonomic Society, 15(3), 147-149.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1597-1611.

Baker, A.G., Berbrier, M.W., & Vallée-Tourangeau, F. (1989). Judgements of a 2 x 2 contingency table: Sequential processing and the learning curve. Quarterly Journal of Experimental Psychology, 41B(1), 65-97.

Baker, A.G., Mercier, P., Vallée-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: The presence of a strong causal factor may reduce judgments of a weaker one. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(2), 414-432.

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

Bar-Hillel, M. (1983). The base rate fallacy controversy. In R. Scholz (Ed.), Decision making under uncertainty. (pp. 39-61). North-Holland: Elsevier Science Publishers.

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

Bar-Hillel, M. (1990). Back to base rates. In R. M. Hogarth (Ed.), Insights in decision making: A tribute to Hillel J. Einhorn. (pp. 200-223). Chicago: University of Chicago Press.

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

Brase, G.L., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. Journal of Experimental Psychology: General, 127(1), 3-21.

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

Cheng, Y. (1995). Covariation detection biases in sufficient and necessary situations. Unpublished doctoral dissertation, University of Ottawa, Ottawa, Ontario, Canada.

Christensen-Szalanski, J.J.J., & Beach, L.R. (1982). Experience and the base rate fallacy. Organizational Behavior and Human Performance, 29, 270-278.

Christensen-Szalanski, J.J.J., & Bushyhead, J. (1981). Physicians' use of probabilistic information in a real clinical setting. Journal of Experimental Psychology: Human Perception and Performance, 7(4), 928-935.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press, Inc.

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

Dellarosa, D. (1988). A history of thinking. In R. J. Sternberg & E. E. Smith (Eds.), The psychology of human thought. (pp. 1-18). Cambridge: Cambridge University Press.

Dickinson, A., Shanks, D.R., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. Quarterly Journal of Experimental Psychology, 36A, 29-50.

Gigerenzer, G. (1991a). From tools to theories: A heuristic of discovery in cognitive psychology. Psychological Review, 98(2), 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, Volume 2. (pp. 83-115). Chichester, England: John Wiley & Sons.

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

Gigerenzer, G. (1993b). The Superego, the Ego, and the Id in statistical reasoning. In G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues. (pp. 311-339). Hillsdale, NJ: Lawrence Erlbaum.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vise versa). In G. Wright & P. Ayton (Eds.), Subjective probability. (pp. 129-161). Chichester, England: John Wiley & Sons.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). Psychological Review, 103(3), 592-596.

Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. D. Cummins & C. Allen (Eds.), The evolution of mind. (pp. 9-29). Oxford: Oxford University Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. Psychological Review, 102(4), 684-704.

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

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. Journal of Personality and Social Psychology, 52(3), 464-474.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and semantics: Volume 3, Speech acts. (pp. 41-58). New York: Academic Press.

Hasher, L., & Zacks, R.T. (1979). Automatic and effortful processes in memory. Journal of Experimental Psychology: General, 108(3), 356-388.

Hasher, L., & Zacks, R.T. (1984). Automatic processing of fundamental information. American Psychologist, 39(12), 1372-1388.

Hertwig, R., & Chase, V.M. Many reasons or just one: How response mode affects reasoning in the conjunction problem. (un pub)

Hertwig, R., & Gigerenzer, G. The "Conjunction Fallacy" revisited: Polysemy, conversational maxims, and frequency judgments. (un pub)

Hilton, D.J. (1990). Conversational processes and causal explanation. Psychological Bulletin, 107(1), 65-81.

Hilton, D.J. (1991). A conversational model of causal explanation. In W. Stroebe & M. Hewstone (Eds.), European Review of Psychology, 2. (pp. 51-81). Chichester, England: John Wiley & Sons.

Hilton, D.J. (1995a). Logic and language in causal explanation. In D. Sperber, D. Premack, & A. J. Premack (Eds.), Causal Cognition: A Multidisciplinary debate. (pp. 495-525). Oxford: Clarendon Press.

Hilton, D.J. (1995b). The social context of reasoning: Conversational inference and rational judgment. Psychological Bulletin, 118(2), 248-271.

Hilton, J., & Fein, S. (1989). The role of typical diagnosticity in stereotype-based judgments. Journal of Personality and Social Psychology, 57(2), 201-211.

Jeffereys, H. (1961). The theory of probability (Third Edition). London: Oxford University Press

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

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. (1982a). On the psychology of prediction. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. (pp. 48-68). Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982b). Variants of uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. (pp. 509-520). Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. Psychological Review, 103(3), 582-591.

Kao, S.-F., & Wasserman, E.A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(6), 1363-1386.

Kirk, R.E. (1995). Experimental design: Procedure for the behavioral sciences (Third Edition). Pacific Grove, CA: Brooks/Cole.

Kleiter, G.D., Krebs, M., Doherty, M.E., Garavan, H., Chadwick, R., & Brake, G. (1997). Do subjects understand base rates? Organizational Behavior and Human Decision Processes, 72(1), 25-61.

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

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? Organizational Behavior and Human Performance, 20, 159-183.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. Journal of Personality and Social Psychology, 39(5), 821-831.

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. Journal of Experimental Social Psychology, 18, 23-42.

Macchi, L. (1995). Pragmatic aspects of the base-rate fallacy. Quarterly Journal of Experimental Psychology, 48A(1), 188-207.

Medin, D., & Bettger, J. (1991). Sensitivity to changes in base-rate information. American Journal of Psychology, 104(3), 311-332.

Medin, D., & Edelson, S. (1988). Problem structure and the use of base-rate information. Journal of Experimental Psychology: General, 117(1), 68-85.

Nisbett, R.E., & Borgida, E. (1975). Attribution and the psychology of prediction. Journal of Personality and Social Psychology, 32(5), 932-943.

Nisbett, R.E., Borgida, E., Crandall, R., & Reed, H. (1976). Popular induction: Information is not necessarily informative. In J. S. Carroll & J. W. Payne (Eds.), Cognition and social behaviour.Hillsdale, NJ: Erlbaum.

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

Nisbett, R.E., Zukier, H., & Lemley, R. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. Cognitive Psychology, 13, 248-277.

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

Schwarz, N., Strack, F., Hilton, D.J., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. Social Cognition, 9(1), 67-84.

Shanks, D.R. (1985). Continuous monitoring of human contingency judgment across trials. Memory and Cognition, 13(2), 158-167.

Shanks, D.R., & Dickinson, A. (1987). Associative accounts of causality judgment. Psychology of Learning and Motivation, 21, 229-261.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture . (pp. 19-136). Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1971).Belief in the law of small numbers. Psychological Bulletin, 76(2), 105-110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), Progress in Social Psychology, Vol 1. (pp. 49-72). Hillsdale, NJ: Lawrence Erlbaum.

Tversky, A., & Kahneman, D. (1982a). Availability: A heuristic for judging frequency and probability. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. (pp. 163-178). Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982b). Causal schemas in judgments under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. (pp. 117-128). Cambridge: Cambridge University Press.

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

Tversky, A., & Kahneman, D. (1982e). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. (pp. 3-20). Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982f). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment 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(4), 293-315.

Wasserman, E.A., Domer, W.W., & Kao, S.-F. (1990). Contributions of specific cell information to judgments of intervent contingency. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16(3), 509-521.

Wright, G. (1984). Behavioral decision theory: An introduction. London: Sage.

Zuckerman, M. (1978). Use of consensus information in prediction behaviour. Journal of Experimental Social Psychology, 14, 163-171.

Zukier, H. (1982). The dilution effect: The role of the correlation and the dispersion of predictor variables in the use of nondiagnostic information. Journal of Personality and Social Psychology, 43(6), 1163-1174.