THE FREQUENCY PRESENTATION EFFECT: DETERMINING ITS NATURE AND CONSISTENCY

by

Tracy L. Lindberg

B.A. (Honours), Simon Fraser University, 1992

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the

Department of Psychology

SIMON FRASER UNIVERSITY

August 1995

ISBN 0-612-06710-6

APPROVAL

Name: Tracy L. Lindberg

Degree: Master of Arts

Title of Thesis: The Frequency Presentation Effect: Determining Its Nature and Consistency

Examining Committee:

Chair: Dr. Vito Modigliani

Dr. Charles Crawford
Senior Supervisor
Professor of Psychology
Department of Psychology
Simon Fraser University

Dr. Cathy McFarland
Associate Professor of Psychology
Department of Psychology
Simon Fraser University

Dr. Rod Wong
External Examiner
Professor of Psychology
Department of Psychology
University of British Columbia

Date Approved: 9 August 1995

Title of Thesis/Project/Extended Essay

The Frequency Presentation Effect: Determining its Nature and Consistency

Author:
(signature)

Tracy Lindberg
(name)

95/8/10
(date)

Abstract

This research replicates and extends the recent work of Cosmides and Tooby (in press) on probabilistic reasoning in judgments under uncertainty. Using a problem well-known for eliciting base-rate neglect, Cosmides and Tooby concluded that subjects were good intuitive statisticians when the information was presented in frequency form. Four experiments were designed as a three-part investigation into the frequency presentation effect reported by Cosmides and Tooby. The goals of the investigation were as follows: (1) to test the replicability of the frequency effect; (2) to isolate the effect of the frequency presentation from the effects of confounding variables; and (3) to re-examine how well subjects' reasoning reflects aspects of a calculus of probability, such as Bayes' Theorem. Results from Experiments 1 and 2 replicated the frequency effect, however the effect was inconsistent with clarified versions of the problem. The effect disappeared in Experiments 3 and 4, once aspects of mathematical difficulty were controlled. In addition, examination of subjects' descriptions of their solutions revealed poor understanding of Bayes' Theorem. These results are discussed in relation to recent research on cognitive biases.

Acknowledgements

I wish to express my thanks to Charles Crawford and Cathy McFarland for their invaluable guidance throughout the duration of this work. I am grateful to Sally Walters, Maria Janicki, and Judith Anderson for the numerous questions they have entertained, and the feedback they have provided. I greatly appreciate the hours of statistical assistance donated by Michael Maraun. Special thanks are directed to Debra Quinn and Bryan Kraushaar, for their respective efforts in the daunting task of coding the questionnaires. I appreciate your dedication. Thanks to Timmy and Kalilah, for their calming influence. Finally, I would like to express my thanks to everyone who donated their thoughts and their time, and who, in varied ways, contributed to the completion of this work.

Table of Contents

List of Tables

Introduction

Research in social cognition, particularly that dealing with judgment under uncertainty, has established that human reasoning is riddled with cognitive biases and fallacies, such as base-rate neglect, representativeness, and the conjunction effect (for example, Kahneman, Slovic & Tversky, 1982), leading researchers to conclude that humans are poor intuitive statisticians. However, this conclusion has recently been challenged. Gigerenzer (1994) and Cosmides and Tooby (in press) argue that studies establishing these biases are based on a Bayesian interpretation of probability, where probability refers to a subjective degree of confidence. From a Bayesian perspective, the probability of a single event can be calculated (Gigerenzer, Swijtink, Porter, Daston, Beatty, & Krüger, 1989), which is what subjects in judgment under uncertainty experiments are usually asked to do. For example, in the Cab Problem (Kahneman & Tversky, 1972), subjects are given base-rate information about the percentage of cabs in different companies, and information about the content and probable accuracy of an eye-witness report, and then asked to judge the probability that the cab involved in the accident was from a particular company. Most subjects answer incorrectly, and researchers have attributed these errors to biases in reasoning, such as the tendency to underutilize the base-rate information (Bar-Hillel, 1980).

However, Cosmides and Tooby (in press) suggest that from an evolutionary perspective, one would expect cognitive mechanisms to be frequentist in nature, designed to take frequency information as input and produce frequencies as output. Based on the findings of Gigerenzer (1991a) that the human mind represents probabilistic information as frequencies for many domains, Cosmides and Tooby apply this frequentist hypothesis to human statistical reasoning. They argue that humans are good intuitive statisticians; some of our inductive reasoning mechanisms embody aspects of a calculus of probability, specifically a frequentist perspective of probability.

From the frequentist perspective, probability refers to the relative frequency with which an event occurs, defined over a specific reference class (Gigerenzer et al., 1989). Because probability is a relative frequency, the probability of a single event, which either happens or does not happen, cannot be calculated. Instead, one can calculate the relative frequency with which that event would happen on average. From this perspective, the single event probability question that is the basic dependent variable in most judgment under uncertainty research is inappropriate. Therefore, these studies are flawed (Gigerenzer, 1994). Subjects' inability to correctly calculate the single event probability in question may be a reflection of the incongruence of this question with the frequentist perspective, rather that a deficit in reasoning.

To test the hypothesis that the mind is a good frequentist statistician, Cosmides and Tooby revisited the Medical Diagnosis Problem, a problem well-known for eliciting base-rate neglect. This problem was originally posed to attending physicians, fourth-year medical students, and house officers (residents) at four Harvard Medical School teaching hospitals by Casscells, Schoenberger, and Grayboys (1978) in the following form:

> If a test to detect disease whose prevalence is 1/1000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? (p. 999)

The problem can be solved formally by using Bayes' Theorem, which provides a rule for revising a base rate (prior probability) based on observed data, or through an informal understanding of the components contained in Bayes' Theorem. In the formal terms of Bayes' theorem, one could calculate the probability of having the disease given a positive result [P(Disease/Positive)] as follows:

$$= \frac{P(\text{Disease}) * P(\text{Positive/Disease})}{[P(\text{Disease}) * P(\text{Positive/Disease})] + [P(\text{Not Disease}) * P(\text{Positive/Not Disease})]}$$

$$= \frac{(1/1000) * 100\%}{[(1/1000) * 100\%] + [(999/1000) * 5\%]}$$

$$= \quad 0.0196 \text{ or } 1.96\%$$

Informally, as explained by Casscells et al.:

> Common sense alone is needed to understand that only one of 1000 people studied will, on the average, have the disease, and 5 per cent of others (0.05 X 999), or roughly 50 persons, will yield (falsely) positive results. Thus only one of 51 positive results will be truly positive, and the chance that any one positive result represents a person with the disease is one in 51, or less that 2 per cent. (1978, pp. 999 - 1000)

Casscells et al. found that only 18% of their subjects provided the correct answer to this problem. Even lower rates might be predicted for less statistically knowledgeable subjects. However, the problem used by Casscells et al. contained information in percentage form, and required subjects to calculate a single event probability, consistent with a Bayesian probability perspective. Based on the frequentist hypothesis, Cosmides and Tooby hypothesized that a frequentist version of this problem, which provided the information in the problem in frequency form and required subjects to calculate a relative frequency answer, might enable subjects to solve the problem correctly. Specifically, they made the following predictions:

> 1. Inductive reasoning performance will differ depending on whether subjects are asked to judge a frequency or the probability of a single event.
> 2. Performance of frequentist versions of problems will be superior to non-frequentist versions.
> 3. The more subjects can be mobilized to form a frequentist representation, the better performance will be.
> 4. Performance on frequentist problems will follow aspects of a calculus of probability, such as Bayes's rule. This is because some inductive reasoning mechanisms do embody aspects of a calculus of probability.
>
> (Cosmides & Tooby, in press, manuscript p. 17-18)

Using various frequentist versions ot this problem, Cosmides and Tooby were able to elicit the correct response from 56 - 92% of their subjects. These frequentist versions, which presented the information in frequency form instead of percentages, also contained several modifications designed to clarify the information in the problem. These modifications included providing the true positive rate (which would have to be assumed in the original version), defining the term 'false positive', and specifying a random sample. The lower end of this range of accuracy resulted when subjects were given clarified versions of the problem in which the information was presented in both frequency and percentage form. The highest accuracy rates were obtained when, in addition to clarifying the information in the problem and presenting the information in only frequency form, subjects were also instructed to actively construct a pictorial representation of the information in the problem. These findings supported their predictions.

After establishing that the presentation of the information in the problem did affect accuracy rates, Cosmides and Tooby investigated whether any aspects of the problem, other than the presentation, could produce the same level of increased accuracy. Examining the effect of the specific modifications introduced to clarify the problem, they concluded that although manipulating aspects of the problem other than the presentation (e.g., defining 'false positive') did result in slight increases in accuracy (increasing accuracy rates from 12% to 36%), the effect of presentation was much more dramatic. Analyzing the presentation effect, they discovered that asking for the answer as a relative frequency produced a larger effect on accuracy rates than presenting the information as frequencies, although both effects were significant.

Given the dramatic effect that changing the presentation to frequency form had on accuracy rates with the Medical Diagnosis problem, these results are worthy of further investigation. If the presentation effect can be substantiated through replication, it has implications for much of the past research establishing base-rate neglect. It would also add to similar research on other cognitive biases in reasoning (see Gigerenzer, 1994). Is the same frequency information effect found with other problems common to the judgment under uncertainty literature, such as the Cab Problem (Kahneman & Tversky, 1972) and social cognition problems (Nisbett & Borgida, 1975; Nisbett, Krantz, Jepson, & Kunda, 1983)? Does the frequentist approach eradicate the cognitive biases in reasoning reported by past research?

However, there may be some cause to suspect that these findings, which support the frequentist reasoning hypothesis, may not be as straightforward or conclusive as they appear. In addition to attempting to replicate the findings from the Cosmides and Tooby study, this investigation identified several issues pertaining to the nature and consistency of the frequency effect. Specifically, the role of the inverse probability interpretation, which was identified as a common error committed by subjects, and the role of mathematical difficulty as a possible confounding variable, were addressed.

In Experiment 6 of their study, Cosmides and Tooby showed that a substantial proportion of the inaccurate responses by subjects receiving the percent information version of the problem were the result of the subjects incorrectly interpreting the false positive rate as an inverse probability. In probability terms, the false positive rate is P(Positive/Not Disease); that is, it is the probability of testing positive given that you do not have the disease. The inverse probability would be P(Not Disease/Positive), or the probability of not having the disease given that you have tested positive (Gigerenzer, 1993). Thus, in the Medical Diagnosis Problem the correct interpretation of the 5% false positive rate is that '5 out of every 100 healthy people will test positive for the disease'. However, if subjects are interpreting the false positive rate as an inverse probability they would be taking it to mean that 5 out of every 100 people who test positive for the disease will actually be healthy (and therefore be falsely positive). Misinterpreting the false positive rate in this way, the answer to the problem becomes 95%, that is the remaining 95 people out of the 100 positives will have the disease. The wording of the false positive information in the frequency version of the problem does not allow for this type of misinterpretation. Therefore, it is important to separate out the effect of the frequency information from the effect of removing the inverse probability interpretation. In Experiment 5, Cosmides and Tooby improved the wording of the percent version of the problem so that the chances of misinterpretation were reduced and, not surprisingly, found fewer inverse probability interpretations.

However, the chances of misinterpretation can only be minimized by holding the wording of the false positive information completely parallel in both the percent and frequency versions. This adjustment was included in the clarified percent version of the problem used for this replication.

In addition, it is possible that the higher accuracy rates generated by the frequency version are the result of something other than the method of presenting the information. The effects of variables that varied concurrently with the frequency presentation manipulation in the Cosmides and Tooby versions of the problem may have invisibly added to the frequency effect, causing the frequency effect to appear stronger and more dramatic than it really is. One such confounding variable examined in this investigation is the level of mathematical difficulty in the problem. This was examined with respect to both the inclusion of a common denominator amongst the rates provided in the problem, and the complexity of the mathematical operations subjects needed to perform to obtain the correct solution.

In the Cosmides and Tooby study, subjects in the frequency conditions were provided with a common denominator of 1000 for the base rate, false positive rate, and sample size. However, subjects in the percent conditions received a denominator of 1000 for the base rate and a percentage, implicitly out of 100, for the false positive rate. This may have required them to do more mathematical manipulation than subjects in the frequency conditions. Thus, it is possible that at least part of the frequency effect observed is actually the effect of simplified calculation requirements resulting from the common denominator in the information presented. To remove this possibility, it is necessary to separate the occurrence of a common denominator in the information provided in the problem from the presentation of the information in frequency form. This was the purpose of Experiment 3.

Furthermore, the percent and frequency versions may be unequal with regard to mathematical difficulty in terms of the mathematical operations which subjects are required to perform. As an illustration of this inequality, in order for subjects in the frequency condition to solve the problem entirely correctly, they must report the probability of having the disease as "1 out of 51". To derive this solution, subjects must correctly identify the groups indicated in the question and engage in addition (adding 1 and 50). By comparison, in order for subjects in the percent condition to solve the problem entirely correctly, they must report the probability as "1.96%". Deriving the latter involves higher level calculations such as division and multiplication of decimal numbers, even if only to translate "1 out of 51" to the equivalent percent form. Therefore, it is possible that the difference in difficulty of mathematical operations required between the two conditions is contributing to the observed frequency effect.

It should be noted that Cosmides and Tooby designated 2% as the correct answer, not the 1.96% mentioned above. However, rather than eliminate the above problem, this introduces another: subjects may not be solving the problem correctly, in terms of the Bayes' theorem standard. Because Cosmides and Tooby were loose in their criteria for a "correct" answer, answers that were correct according to a strict Bayes' theorem standard were grouped together with answers that only approximated this standard. As a result, it is impossible to distinguish whether subjects identified by Cosmides and Tooby as showing Bayesian reasoning are incorporating all aspects of Bayes' theorem in their reasoning or overlooking significant elements. To illustrate, subjects who answered "1 out of 50" would be identified by Cosmides and Tooby as having solved the problem correctly, when in reality they may have failed to identify that the individual who has the disease must be added to the group of 50 individuals who are *healthy* and test positive in order to determine the total number of positive test results. In terms of Bayes' theorem, such individuals are failing to add P(Disease)*P(Positive/Disease) into the denominator. This has direct bearing on conclusions concerning the fourth prediction made by Cosmides and Tooby, that subjects' reasoning mechanisms and performance embody aspects of a calculus of probability.

Although omitting the 'diseased and positive' individual from the total number of positive cases does not produce a noticeable difference in the answer for this problem, the resulting error is much more visible, and relevant, in problems where the base rate is greater. In addition, because of the particular rates given in the problem, it is also not apparent whether subjects are making another error: applying the false positive rate to the total sample of sick and healthy individuals (1000), instead of the sample of healthy individuals only (999 of the 1000). Therefore, in the final part of this investigation the base rate and false positive rate provided in the problem were modified so that (1) the mathematical operations required by the subjects were approximately equivalent in difficulty for both frequency and percent versions, and (2) the correct Bayes' theorem answers could be more easily distinguished from approximations which did not reflect an understanding of all of the components inherent in the Bayes' theorem solution.

In sum, the following experiments were designed as a three-part investigation into the frequency presentation effect reported by Cosmides and Tooby. The goals of the investigation were as follows: (1) to test the replicability of the frequency effect; (2) to isolate the effect of the frequency presentation from the effects of potential confounding variables; and (3) to re-examine how well subjects' reasoning reflects aspects of a calculus of probability, such as Bayes' theorem.

Experiment 1

Experiment 1 was designed as a modified replication of the 'presentation' and 'clarification' manipulations. Several changes were incorporated to provide a clearer examination of these effects. Firstly, the Cosmides and Tooby study did not examine accuracy rates for a frequency version of the *original* problem, without any added clarifications of wording or definition of terms. This comparison was included in this replication. In addition, an examination of the effect of presenting information in decimal form was also added. Finally, some minor inconsistencies in wording between the percent and frequency conditions in the Cosmides and Tooby study were identified and eliminated. This included controlling the wording of the false positive rate information to minimize the inverse probability misinterpretation.

Method

Participants

One hundred twenty-six male and female undergraduate students at Simon Fraser University participated in Experiment 1, 21 in each of six conditions. Subjects partially fulfilled a requirement for a psychology course by participating.

Procedure

A 2X3 design was utilized, with 2 levels of the Clarification variable (Original and Clarified) and 3 levels of the Presentation variable (Percent, Decimal Probability, and Frequency). The Decimal Probability information condition was not present in the Cosmides and Tooby study but was included to provide a test of the equivalency of percent and decimal probability presentations. In addition, the following minor inconsistencies across conditions were identified, and corrected. First, the word "actually" in the problem's question was italicized in the Frequency conditions, but not in the Percent conditions. This was remedied by removing the italics in the Frequency condition, so that "actually" was in plain font for all conditions. Similarly, some frequency versions of the problem contained an explicit definition of the sample size as 1000. This reference was not present in the percent versions, and was shown by Cosmides and Tooby to have no effect in the frequency conditions. Therefore this reference to the sample size was omitted from all conditions. Thus, the content of the problem was held completely constant across conditions, except for the manipulations specified below.

The following example of the Original and Clarified Frequency conditions illustrates the clarification manipulation. The wording of the problem presented to subjects in the Original Frequency condition was as follows:

> If a test to detect a disease whose prevalence is 1 out of 1000 has a false positive rate of 50 out of 1000, on average, how many people who test positive for the disease will actually have the disease, assuming you know nothing about their symptoms or signs?
>
> _______ out of _________

For the Clarified Frequency condition the problem was worded as follows:

> 1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, for every 1000 people who are perfectly healthy, 50 of them will test positive for the disease.
>
> Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of these people.
>
> Given the information above:
> on average,

How many people who test positive for the disease will actually have the disease?

_______ out of ________

The modifications to the original problem present in the clarified version were identical to those in the Cosmides and Tooby study, with the exception of the changes noted above. For the presentation manipulation, in the original versions of the problem the prevalence rate was presented as 1 out of 1000 for the Frequency condition (see above), 1/1000 for the Percent condition (to be consistent with the Cosmides and Tooby and Casscells et al. wording), and 0.001 for the Decimal Probability condition. The false positive rate was presented as 50 out of 1000, 5%, and 0.05, respectively.

In the clarified versions the differences between conditions in terms of presentation were more substantial. In the first sentence, the prevalence rate was presented as "1 out of every 1000" for the Frequency condition, "1/1000" for the percent condition, and "The rate of disease X among Americans is 0.001" for the Decimal Probability condition. The last sentence of the first paragraph presented the false positive rate. For the Frequency condition it read "for every 1000 people who are perfectly healthy, 50 of them will test positive for the disease." For the Percent condition, the sentence read "of all people who are perfectly healthy, 5% will test positive for the disease," and for the decimal probability condition it read "for all people who are perfectly healthy, the probability of testing positive for the disease is 0.05." In addition, the conditions varied in terms of the wording of the question posed to the subjects. In the Frequency condition, the question read "How many people who test positive for the disease will actually have the disease? ____out of ______ ". In the Percent and Decimal Probability conditions, subjects were asked "What is the chance that a person found to have a positive result actually has the disease?" and directed to provide the answer as a percentage or decimal probability, respectively. In the original versions of the problem, the same variations in the question posed were included in the main text of the problem (see example above). Copies of all questionnaires are provided in Appendix A.

Subjects were given as much time as they required to complete the problem.

Results

An alpha level of 0.05 was used for all statistical tests. Logistic regression was utilized to test for significant differences in accuracy between conditions. This procedure allows for the same comparisons on a dichotomous variable that are available for a continuous variable (Lunneborg, 1994). The effect of the presentation manipulation was analyzed for each clarification level. The results provided only a partial replication of Cosmides & Tooby's findings (see Table 1). When subjects were solving the Original version of the Medical Diagnosis Problem, there was a significant effect for presentation. A higher proportion of correct responses was elicited in the Frequency presentation condition (47.6%) than in either the Percent (0%) or Decimal Probability conditions (4.8% , $p=0.0325$). Unexpectedly, there was no effect for presentation when subjects were solving the Clarified version of the problem. The proportion of correct answers was 28.6% for both the Percent and Frequency conditions, and 19.0% for the Decimal Probability condition. These rates were not significantly different ($p=0.7183$). There was no significant difference between the accuracy rates for the Percent and Decimal Probability conditions ($p=0.7333$).

Discussion

Although the frequency information had the predicted effect when subjects were presented with the Original (unclarified) version of the problem, this effect did not emerge once the information in the problem was clarified. This finding contradicts the result reported by Cosmides and Tooby, who found an effect for presentation beyond the effect of clarifications to the wording of the problem. Therefore, the clarified problem comparisons were repeated, to test if this result would replicate. In addition, because there was no difference found between the percent and decimal probability presentations, the Decimal Probability condition was not included in further experiments.

Experiment 2

Experiment 2 was a direct replication of the Clarified Percent and Frequency conditions in Experiment 1. In addition, because the accuracy rates appeared low, relative to the parallel rates reported by Cosmides and Tooby, the Clarified Frequency version of the problem used by Cosmides and Tooby was included to provide a direct comparison with the results from their study. As this latter version had generated accuracy rates of 72% - 80% in their sample of Stanford undergraduates, it was predicted that any difference in accuracy rates from the Clarified Frequency version from Experiment 1 would be in the direction of greater accuracy.

Method

Participants

The participants were 78 male and female undergraduate students at Simon Fraser University, 26 in each of three conditions. Participation in the

Table 1

Experiment 1: Percentage of Correct Responses by Presentation for Original and Clarified Problems

	Problem	
Presentation	Original	Clarified
Percent	0.0	28.6
Decimal Probability	4.8	19.0
Frequency	47.6	28.6
p	0.0325*	0.7183

Note. *n*= 21 per group.

experiment partially fulfilled a requirement for a psychology course.

Procedure

There were three clarified versions of the Medical Diagnosis problem given to subjects. Subjects in the first condition received the Clarified Percent version from Experiment 1. Subjects in the second condition received the Clarified Frequency version from Experiment 1, which had been slightly modified from the Cosmides and Tooby version to correct for discrepancies in wording from the percent version. Subjects in the third condition received the Clarified Frequency version of the problem used by Cosmides and Tooby (see Appendix A).

As in Experiment 1, subjects were given as much time as necessary to solve the problem.

Results

As in Experiment 1, logistic regression was used to compare differences in accuracy rates among the three conditions. A higher accuracy rate was obtained for the modified Clarified Frequency condition in this replication (50.0%) than in Experiment 1, bringing it more in line with the 47.6% accuracy rate in the Original Frequency condition in Experiment 1. The accuracy rate for the Percent condition was 23.1%; this difference was in the predicted direction and, unlike the clarified version results from Experiment 1, represents a significant effect for presentation ($p=0.0479$). However, the unmodified Clarified Frequency condition, taken directly from the Cosmides and Tooby study, failed to elicit higher accuracy rates; the percentage of correct responses was identical to that of the Percent condition (23.1%), and significantly lower than that of the modified Frequency condition ($p=0.0479$).

Discussion

The results from Experiment 2 replicated the significant effect of presentation on accuracy rates reported by Cosmides and Tooby for the clarified versions of the problem used in Experiment 1. However, the increase in accuracy rates for the Frequency condition was not as dramatic as the 72 - 80% accuracy rates achieved by Stanford subjects in the Clarified Frequency condition in the Cosmides and Tooby study. The highest accuracy rate, achieved by subjects in the modified Frequency condition, was 50%. In addition, contrary to what was predicted, subjects did worse (23.1%) when presented with the unmodified Cosmides and Tooby Frequency version. Thus, comparing the performance on this frequency version with the percent version no presentation effect emerged. Taken together with the lack of a presentation effect for the clarified versions of the problem in Experiment 1, this suggests that the frequency effect is less consistent when clarifications are provided within the problem.

Experiments 1 and 2 provided evidence that the frequency effect can be replicated, most clearly when the original wording of the problem was used. The effect was inconsistent when the wording of the information in the problem was clarified. Experiments 3 and 4 were intended to investigate whether the effect observed for the frequency presentation is confounded with mathematical difficulty, and to examine how well subjects' reasoning reflects the components of Bayes' theorem.

Experiment 3

Experiment 3 was designed to separate the occurrence of a common denominator in the information provided in the problem from the presentation of the information in frequency form. In the Cosmides and Tooby study, the base rate was always given as 1/1000 (or 1 out of every 1000). For frequency versions of the problem, the false positive rate was given as 50 out of 1000, whereas for percent versions of the problem, it was given as 5%, implicitly out of 100. Therefore, unlike subjects in the frequency conditions, subjects in the percent conditions did not receive the information in common denominator form. This study varied the occurrence of the common denominator independent from the frequency presentation.

Method

Participants

The participants were 130 male and female undergraduate students at Simon Fraser University, 26 in each of five conditions. For all subjects, participation in the experiment partially fulfilled a requirement for a psychology course.

Procedure

A 2X2 design was utilized to vary the presence or absence of a common denominator along with the form of presentation (Frequency or Percent). This design contained the problems from the Original Frequency presentation and Original Percent presentation conditions run in Experiment 1 as the Common Denominator Frequency presentation condition and the No Common Denominator Percent presentation condition, respectively. Two additional conditions, the Common Denominator Percent presentation and the No Common Denominator Frequency presentation, were included to complete the design. In order to present a common denominator with a percent form false positive rate, it was necessary to change the base rate to a denominator of 100. To provide a direct comparison, the same change was made to the frequency version (see

Appendix A for questionnaires). As a result, there were two different versions of the problem in the Common Denominator Frequency condition: one with a base rate of "1 out of 1000" and the other with a base rate of "1 out of 100". The No Common Denominator conditions both contained a base rate of 1/1000. Changing the base rate to 1/100 also changed the solution to the problem, and resulted in a slightly greater ability to distinguish exactly correct answers (1/6 or 0.167) from approximations (1/5 or 0.2). This provided a preliminary examination of whether changing the base rate would give insight into the correctness of subjects' reasoning. However, the correctness of subjects' reasoning was examined more thoroughly in Experiment 4.

Results

As with Experiments 1 & 2, logistic regression was used to test for differences in accuracy between conditions (see Table 2). Overall, there was no significant effect for the presence of the common denominator, regardless of whether the denominator in the frequency presentation was 100 or 1000. For subjects in the No Common Denominator condition, presenting the information in frequency form did not produce significantly higher rates of accuracy compared to percent information (15.4% vs. 3.8%, $p=0.1901$).

A more complex analysis of the presentation effect was possible for subjects in the Common Denominator condition. Comparing accuracy rates by presentation when the base rate given in both the frequency and percent versions was 1/100, there was a significant presentation effect as predicted; higher accuracy rates were elicited when the information was in frequency form (34.6%) than when it was in percent form (11.5%) ($p=0.0329$). However, as mentioned above, with this base rate exactly correct answers were more distinct from approximations than with the 1/1000 base rate. Once the standard of correctness

Table 2

Experiment 3: Percentage of Correct Responses by Presentation for Problems With and Without a Common Denominator

	Occurrence of Common Denominator		
	Yes		No
Presentation	1/100	1/1000	1/1000
Percent			
Exact and Approximate	11.5[ac]		3.8
Exact Only	3.8[b]		
Frequency			
Exact and Approximate	34.6[a]	7.7[c]	15.4
Exact Only	15.4[b]		
p	0.0329*[a]	0.6402[c]	0.1901
	0.7331[b]		

Note. n = 26 per group.

[a]Comparing the percentages of correct answers including approximations for Percent and Frequency presentations.

[b]Comparing the percentages of exact correct answers only for Percent and Frequency presentations.

[c]Comparing the percentage of correct answers for the Percent presentation with a base rate of 1/100 and the percentage of correct answers for the Frequency presentation with a base rate of 1/1000.

was narrowed to allow for exact answers only, accuracy rates dropped to 15.4% for subjects in the Frequency condition, and 3.8% for subjects in the Percent condition. With this standard, there was no effect for the frequency presentation ($p=0.7331$). Subjects in the Common Denominator Frequency condition who received the version of the problem with a base rate of 1/1000 showed lower accuracy rates; only 7.7% of subjects in this condition reported the exact or approximate answer. Again, comparing this rate to the percent version, there was no significant effect for presentation ($p=0.6402$).

Discussion

The results from Experiment 3 did not provide evidence of a consistent frequency effect once the occurrence of a common denominator in the problem was controlled. Two comparisons of the effect of the frequency presentation when the problem contained a common denominator yielded contradictory results; there was a significant effect when the base rate provided in the Frequency condition was 1/100, but no significant effect when the base rate was 1/1000. In addition, there was no significant effect for frequency presentation when the problem did not contain a common denominator. Thus, the effect of the frequency presentation, once the common denominator variable is controlled, appears to be inconsistent at best, and generally nonsignificant.

In addition, closer examination of the responses for subjects who received the 1/100 base rate suggested that some subjects were obtaining approximately correct answers without including all components of Bayes' theorem in their reasoning. Specifically, these subjects were reporting the answer as "1 out of 5" or "20%" instead of "1 out of 6" or "17%". Once the standard for a correct answer was narrowed to include only exact answers or 1/6 or 17%, the frequency effect in the common denominator conditions with the 1/100 base rate disappeared. This preliminary finding suggested some subjects were being incorrectly classified as reporting the correct answer under the standard used in the Cosmides and Tooby study, and that the accuracy of subjects' reasoning and responses deserved more thorough examination. Experiment 4 was designed to accomplish this goal.

Experiment 4

Experiment 4 was designed to allow for an examination of the effect of the frequency presentation once the possible confounding variable of mathematical difficulty was removed, and to provide a more complex analysis of the correctness of subjects' reasoning. This was accomplished by modifying the base rate and false positive rate in the problem. In addition, the true positive rate, which was 100% in the problem used by Cosmides and Tooby, was varied so that it was possible to identify whether subjects correctly included this information in their reasoning.

The rates provided in the problem were chosen in light of the following constraints: (1) the mathematical operations required for correctly solving the problem were approximately equal in difficulty for frequency and percent presentations; (2) the false positive rate was not above 50%, and the true positive rate was not less than 50%; (3) the base rate was less than and not equal to 50%; (4) no repetition occurred among the rates provided; and (5) answers that correctly included all aspects of Bayes' theorem were discernible from approximations. Constraints (2) and (3) were included to ensure that the rates would appear reasonable to the subjects, and the likelihood of subjects ignoring the rates would be minimized. Constraint (4) was included to minimize the possibility of subjects confusing one rate with another, and to ensure the rates used could be clearly identified in the calculations and explanations provided by subjects. Unfortunately, fully satisfying constraint (5) while maintaining equivalent ease of mathematical operations and satisfying the other constraints proved difficult. When the true positive rate was 100%, it was possible to choose a base rate and false positive rate that would satisfy all of the constraints. However, when the true positive rate was varied from 100%, rates which satisfied all constraints were difficult to identify. Therefore, for these 2 conditions, the rates were chosen to satisfy constraints (1) through (4) and a coding scheme was developed to identify the correctness of subjects' reasoning through their responses to several additional questions. Although correct responses were automatically discernible from approximations in the 100% true positive versions, this coding scheme was applied to all responses, and provided additional information about the subjects' reasoning.

It was predicted that equating mathematical difficulty would decrease the effect of the frequency presentation. In addition, it was predicted that subjects' reasoning would not show all aspects of Bayes' theorem, and that most subjects would be providing approximate answers only.

Method

Participants

The participants were 100 undergraduate students at Simon Fraser University, 25 subjects in each of four conditions. Subjects' participation in the experiment partially fulfilled a requirement for a psychology course.

Procedure

A 2X2 design was utilized, with 2 levels of Presentation (Percent and Frequency) and 2 levels of True Positive Rate (100% and 80%). All versions of the problem included a base rate and false positive rate specifically selected to provide a similar level of mathematical difficulty between conditions, and to differentiate as much as possible between answers which were correct according to the Bayes' theorem standard from approximate answers that did not contain all Bayes' theorem components. For the versions with a true positive rate of 100%, the base rate was 20% and the false positive rate 50%. For the versions with a true positive rate of 80%, the base rate was 40% and the false positive rate was 30%. Copies of each version are included in Appendix A. Because of the different rates used, the correct answer was 20 out of 60 or 33% in the 100% true positive conditions, and 16 out of 40 or 40% in the 80% true positive rate conditions. As several errors in reasoning could also produce an answer of 40% in the latter conditions, it was necessary to analyze subjects' explanations of their reasoning in order to determine whether the 40% truly reflected a correct answer.

To this end, after solving the problem, subjects were asked to describe how they had arrived at their answer to the problem, and instructed to show any steps or calculations they utilized. This response was later coded by three raters for correctness and completeness of Bayes' Theorem reasoning (see Appendix B for coding protocol). As well, subjects' interpretation of the question was coded from this information. Subjects were also directed to identify the combinations of information that were necessary to consider in solving the problem from a list of 8 alternatives, (A) through (H). Among these alternatives, (B) corresponded to P(Disease) * P(Positive/Disease), the numerator component of Bayes' Theorem, and (D) corresponded with P(Not Disease) * P(Positive/Not Disease), the additional component present in the denominator. The six other alternatives were various possible combinations of the information presented in the problem, but were not components of Bayes' Theorem. Subjects were also provided with the opportunity of reporting other combinations they considered that were not included in the list. This identification of the Bayes' components in the list of alternatives was intended to provide a measure of subjects' abstract understanding of Bayes' Theorem. Subjects were asked to indicate how they used the combinations they identified to arrive at their answer. This response provided another measure of their solution to the problem.

Finally, subjects identified the number of courses they had taken in various fields, including mathematics and statistics, and reported their cumulative grade point average. Subjects also completed the Wonderlic Personnel Test, a 12-minute test of mental aptitude. These measures were examined with respect to their relationship to accuracy rates.

Results

In addition to the coding mentioned above, each rater identified each subject's answer to the problem at three points: (A1) the initial answer provided; (A2) the answer provided when describing how the problem had been solved; and (A3) the answer provided when explaining how the combinations of information identified as necessary for solving the problem were used. These three scores allowed for an examination of whether subjects' reasoning changed after being probed to explain their solution, and again after attempting to identify the necessary combinations of information in a more abstract form. Some subjects failed to provide an answer for A3. In these cases, it was assumed that their reasoning had not changed from the last point and A3 was recorded as the same as A2. In addition, 1 subject provided an answer for A1 only.

Agreement among raters for A1, A2, and A3 was 100%. Over all conditions, 24 out of 99 subjects (24.2%) changed their answer from A1 to A2, with 4 of these 24 answering incorrectly at A1 but correctly at A2, and 1 answering correctly at A1 but incorrectly at A2. Of the 24, 14 were in the percent condition, and 10 were in the frequency condition. From A2 to A3, 8 out of 99 subjects (8.1%) changed their answer, with 1 changing to the correct response, and 1 changing from correct at A2 to incorrect at A3. Of the 8, 5 were in the percent condition and 3 were in the frequency condition.

Logistic regression was used to test for differences in accuracy by presentation for A1, A2, and A3 (see Table 3). Among the answers reported for A1, in the 100% True Positive condition 16% of subjects who received the percent presentation and 20% of subjects who received the frequency presentation provided the correct answer. These rates were not significantly different (p=0.7135). In the 80% True Positive condition, the accuracy rates were 32% for the Percent presentation and 20% for the Frequency presentation. Again, these rates were not significantly different (p=0.3369). The standard of accuracy was stricter for A2, as the subject's description of the answer was used to code whether the subjects had correctly solved the problem. As a result, in the 80% True Positive condition, some responses that were coded as correct at A1 were recognized as errors at A2. Specifically, 2 such cases were identified among the Percent presentation condition, and 4 among the Frequency presentation condition. The accuracy rates recorded for these conditions were 20% and 4%, respectively. These rates were not significantly different (p=0.1149). In the 100% True Positive condition, correct answers were provided by 28% of subjects who received the Percent presentation and 24% of subjects who received the Frequency presentation. These rates were not significantly different (p=0.7473). Examining the responses at A3, where the standard of accuracy was once again loosened, accuracy rates rose slightly, to 32% and 24% with the 100% True Positive rate and 24% and 20% with the 80% True Positive rate. Again there was no significant effect for presentation (p=0.5298 and p=0.7331).

Table 3

Experiment 4: Percentage of Correct Responses at A1, A2 and A3 by Presentation and True Positive Rate

	Answer Used		
Presentation	A1	A2	A3
Percent			
100% True Positive	16	28	32
80% True Positive	32	20	24
Frequency			
100% True Positive	20	24	24
80% True Positive	20	4	20
p	0.7135[a]	0.7473[a]	0.5298[a]
	0.3369[b]	0.1149[b]	0.7331[b]

Note. n = 25 per group.

[a]Comparing across presentation for the 100% True Positive condition.

[b]Comparing across presentation for the 80% True Positive condition.

Another measure of the correctness of subjects' reasoning was provided by rating the level of Bayes' Theorem reasoning reflected in the description of their solution. This variable was coded on a 3-point scale, where 1 indicated complete understanding, 2 indicated partial understanding, and 3 indicated little to no understanding. Agreement between raters on this variable was 95%. In the cases where all 3 raters did not agree, the majority code was used. As well, subjects' abstract understanding through identification of the Bayes' Theorem components was coded on a 5-point scale, ranging from completely correct identification to completely incorrect. Agreement between raters on this variable was 100%. The numbers of subjects coded at each level for these variables is presented by group in Table 4.

The majority of subjects in each condition were coded as showing evidence of little or no Bayes Theorem reasoning, except for the 80% True Positive Rate Frequency condition, in which slightly more subjects showed partial Bayes' Theorem Reasoning. Of the responses of subjects receiving the Percent presentation, 11 showed completely correct reasoning. For the Frequency presentation, 7 responses showed completely correct reasoning. An analysis of variance performed on the Bayes' Theorem reasoning levels yielded a main effect for presentation, with better understanding being evidenced by subjects in the Percent presentations ($p=0.043$) (see Table 5). This was in the opposite direction to the presentation effect hypothesized by Cosmides and Tooby.

Examining the measure of abstract understanding again showed the majority of subjects in each condition were completely incorrect in their identification of the necessary components (see Table 4). Only 6 of the subjects receiving the Percent presentation, and 6 of the subjects receiving the Frequency presentation correctly identified the Bayes' Theorem components and no others. As presented in Table 6, an analysis of variance performed on these scores

Table 4

Subjects at Each Reasoning Level by Presentation and True Positive Rate

	Presentation			
	Percent		Frequency	
Reasoning Level	100% True Positive	80% True Positive	100% True Positive	80% True Positive
Bayes' Theorem Reasoning				
1 = Complete	6	5	6	1
2 = Partial	7	10	3	7
3 = Little to None	12	9	16	17
	(2.24)	(2.17)	(2.40)	(2.64)
Identification of Components				
1 = both components only	0	6	3	3
2 = both components & others	1	2	1	3
3 = one component only	2	4	1	2
4 = one component & others	4	6	6	5
5 = neither component	18	7	14	12
	(4.56)	(3.21)	(4.08)	(3.80)

Note. n = 25 per group, except n = 24 for Percent 80% True Positive on Bayes' Theorem Reasoning. Values in parentheses represent group mean scores.

Table 5

Analysis of Variance for Bayes' Theorem Reasoning Levels

Source	df	F	p
Presentation (P)	1	4.193	0.043*
True Positive (T)	1	0.305	0.582
P X T	1	1.037	0.311
Residual	95	(0.586)	

Note. Value enclosed in parentheses represents a mean square error.

Table 6

Analysis of Variance for Identification of Bayes' Components

Source	df	F	p
	Overall Analysis		
Presentation (P)	1	0.063	0.853
True Positive (T)	1	16.248	0.003*
P X T	1	7.104	0.050*
Residual	95	(1.810)	
	for Percent Presentation		
True Positive	1	13.991	0.000*
Residual	48	(1.557)	
	for Frequency Presentation		
True Positive	1	0.481	0.491
Residual	48	(2.038)	

Note. Values enclosed in parentheses represent mean square errors.

yielded a significant main effect for the true positive rate (p=0.003), and a significant interaction between the true positive rate and presentation (p=0.050). Abstract understanding was better among subjects in the 80% True Positive condition with the Percent presentation (p=0.000), but not with the Frequency presentation (p=0.491).

Raters also identified what question the subject appeared to be answering from the description of the solution. Agreement between raters was 97%. These data are presented in Table 7. Overall, approximately 60% of subjects appeared to be answering the question asked. This ranged from 56% to 65% in various conditions. Among those who appeared to be answering a different question, the most common misinterpretations were calculating the number who had the disease out of the total sample (the base rate) or P(Disease), and calculating the total number testing positive in the sample or P(Positive). This data was missing for one subject, and could not be clearly identified for 13 cases.

The responses of subjects answering incorrectly were examined by one rater to discern any commonly committed errors. These findings are presented in Table 8. Some subjects were coded as committing more than one error. One of the most common errors overall was failing to restrict the application of the false positive rate to healthy individuals only. This error was committed by 15 of the 81 subjects answering incorrectly. In addition, 9 subjects failed to add the individuals who had the disease and tested positive into the total number of positive results. Of these 9, 8 were subjects who received the information in frequency presentation. Another common error was reporting the base rate only (n = 9) or the base rate multiplied by the true positive rate (n = 6). As well, several subjects calculated the total number of positive results for the sample, or applied the base rate to the false positive individuals. Even with the wording of the problem clarified to reduce the likelihood of interpreting the false positive as

Table 7

Subjects' Interpretation of Question by Presentation and True Positive Rate

	Presentation			
	Percent		Frequency	
Question Answered	100% True Positive	80% True Positive	100% True Positive	80% True Positive
Correct Interpretation				
Total With Disease & Positive Out of Total Positives [P(D/+)]	15	16	14	14
Alternative Interpretations				
Total With Disease & Positive Out of Total With Disease [P(+/D)]	0	0	1	1
Total Positives Out of the Total Sample [P(+)]	3	1	2	1
Total With Disease Out of the Total Sample [P(D)]	3	6	0	0
Total Positives Out of Total With Disease [P(+)/P(D)]	3	0	2	0
Total Healthy & Positive Out of Total Sample [P(~D/+)]	0	0	0	1
Unclear	1	2	5	5

Note. n = 25 per group except n = 24 for Frequency 100% True Positive.

D = disease. ~D = healthy. + = positive.

Table 8

Errors Committed by Subjects by Presentation and True Positive Rate

	Presentation			
	Percent		Frequency	
Error	100% True Positive[a]	80% True Positive[b]	100% True Positive[c]	80% True Positive[d]
*Inverse false positive	1	2	0	3
Base rate only	3	0	4	2
Base rate X true positive rate	0	4	1	1
True positive rate	0	2	1	0
False positive rate	0	1	0	1
True positive rate - false positive rate	0	3	0	3
False positive rate +/- base rate	2	0	0	0
False positive rate / true positive rate	0	0	0	1
Diseased - true positive rate	0	0	0	1
*Failed to add diseased and positive individuals to total positives	1	0	6	2
Total positives in sample	3	2	1	1
Total positives / total healthy	0	1	0	0

Continued...

Table 8 (continued)

	Presentation			
	Percent		Frequency	
Error	100% True Positive[a]	80% True Positive[b]	100% True Positive[c]	80% True Positive[d]
Applied base rate to total positives calculated	1	0	0	2
Applied base rate to false positives	1	0	4	2
*Applied false positive to total sample or diseased	5	3	6	1
Unclear, guessing or other	2	2	2	6

*Errors identified *a priori*.

[a]n = 18.

[b]n = 19.

[c]n = 20.

[d]n = 24.

an inverse probability, 6 subjects appeared to make this mistake.

Finally, the relationship between the mental aptitude score and ability to solve the problem was examined. Regardless of which accuracy scores were used, A1, A2 or A3, there was a significant difference in mean mental aptitude score between those reporting the correct response and those reporting an incorrect response (see Table 9). Subjects' scores on the Wonderlic Personnel Test ranged from 13 to 37. Using the strictest standard of accuracy (A2), the mean mental aptitude score was 26.47 for individuals reporting the correct response and 23.83 for individuals answering incorrectly ($p=0.0253$). In addition, subjects reporting the correct answer reported having taken more mathematics and statistics courses on average (4.13) than those answering incorrectly (1.75). Scores on this variable ranged from 0 to 21, with 9 subjects failing to provide a response. Reported cumulative grade point averages ranged from 1.95 to 4.17, with little apparent difference in the averages for those answering correctly (3.02) and those answering incorrectly (2.97). This information was not provided by 3 subjects. Due to the number of missing values, and the unstandardized nature of the responses, information on courses taken and cumulative grade point average was examined for trends only.

Discussion

The results from Experiment 4 supported both predictions. Once mathematical difficulty was controlled by changing the rates provided in the problem, the frequency presentation effect disappeared. In addition, coding of subjects' detailed responses showed that the majority evidenced poor intuitive understanding of Bayes' theorem. Consistent with this, the rates of accuracy among subjects receiving the frequency presentation, which varied from 4 - 24% depending on the standard of correctness used, were much lower than the

Table 9

Mean Mental Aptitude Score, Cumulative Grade Point Average and Mathematics and Statistics Courses Taken for Subjects Answering Correctly and Incorrectly at A1, A2, and A3

	Correctness of Answer		
Answer Examined	Correct	Incorrect	p
	Mental Aptitude Scores		
A1	26.14	23.82	0.0376*
A2	26.47	23.83	0.0253*
A3	26.28	23.68	0.0156*
	Cumulative Grade Point Average		
A1	2.97	2.99	
A2	3.02	2.97	
A3	3.02	2.97	
	Mathematics and Statistics Courses		
A1	4.11	1.65	
A2	4.13	1.75	
A3	3.76	1.69	

accuracy rates reported by Cosmides and Tooby with comparable clarified versions, the average of which was 73.3%. Although the addition of the 80% true positive rate could be expected to increase the difficulty of the problem somewhat, this dramatic difference in accuracy rates was also found for problems with a 100% true positive rate. Given that the rates in the problem were specifically chosen to reduce the confusion between correct answers and incorrect answers that appeared approximately correct, the higher accuracy rates reported by Cosmides and Tooby may be an artifact of the inability to differentiate incorrect approximations. More accurate detection of correct responses, coupled with the removal of variation in mathematical difficulty between the percent and frequency versions appear to have eliminated the presentation effect.

Examination of subjects' coded reasoning levels revealed a presentation effect opposite to that predicted. Judging by the level of Bayes' theorem reasoning reflected in their answers, subjects receiving frequency presentations had a poorer understanding of the solution to the problem than those receiving percent presentations. When examining their ability to correctly identify the abstract components in Bayes' theorem in terms of the information in the problem, understanding was better with the 80% true positive rate for subjects receiving the percent presentation. It may be that the 80% rate made the use of the true positive rate information more salient.

An examination of the errors committed by subjects revealed few overall trends. Although an attempt was made to reduce the possibility of misinterpreting the false positive rate as an inverse probability, this error was still committed by several subjects. However, this occurrence was independent of the presentation received. Two other errors were identified a priori which would have resulted in an apparently correct answer in the Cosmides and Tooby study.

First, subjects could be failing to include the individuals who have the disease and test positive into the group of positives. This error was committed by a number of subjects, most of whom received the frequency presentation. Second, subjects could be failing to restrict the application of the false positive rate to healthy individuals only. Again, this was one of the more common errors found, however it appeared to occur independent of presentation. The inability to detect these errors, especially the first one, may have contributed to the higher accuracy rates, particularly with the frequency presentation, in the Cosmides and Tooby study. As performance on the Medical Diagnosis Problem is considered to be a good example of base rate neglect (Cosmides and Tooby, in press), it is also worth noting that contrary to neglecting the base rate, overattending to the base rate, or conservatism, was one of the more common errors.

Many subjects appeared to have misunderstood the question that was being posed. However, given that this information was inferred from their solution to the problem, it must be viewed with caution. For the majority of responses, it is impossible to determine whether this misunderstanding of the question was the cause or result of the errors in reasoning.

Although no relationship emerged between the accuracy rates and the presentation of the information, the subjects' ability to solve the problem correctly was related to their measured mental aptitude. Subjects who solved the problem correctly had a higher average score on the mental aptitude measure than those solving the problem incorrectly. Along with general ability, specific knowledge of mathematics and statistics, as measured by the number of courses in these subjects taken, was also higher for those reporting correct answers. Thus, it appears that general mental aptitude and knowledge of statistics are more closely related to the ability to correctly solve the problem than manner in which the information is presented.

General Discussion

This investigation was a three-part examination into the nature and consistency of the frequency presentation effect reported by Cosmides and Tooby (in press). Experiments 1 & 2 tested the replicability of the effect. Using the original unclarified version of the Medical Diagnosis Problem used by Casscells et al. (1978) the effect replicated. However, once clarifications were included in the problem, replication of the effect was inconsistent.

Experiments 3 & 4 examined whether the frequency presentation effect would occur once possible confounding aspects of mathematical difficulty were controlled. Experiment 3 manipulated the occurrence of a common denominator among the base rate and false positive rate presented in the problem independent of the presentation of that information as a frequency or percent. When the problem did not have a common denominator no presentation effect was evident. However, when these rates shared a common denominator the presentation effect was inconsistent, occurring when the denominator was 100, but not when it was 1000 or when a stricter standard of accuracy was used.

In Experiment 4, the rates provided in the problem were altered so that the difficulty of mathematical operations would not be dependent on the presentation. At the same time, these rates decreased the possibility of misidentifying incorrect solutions as correct. No presentation effect was found among the accuracy rates. This experiment also provided an opportunity to explicitly examine subjects' reasoning in terms of the understanding of Bayes' Theorem reflected in their solutions. Cosmides and Tooby predicted that subjects' reasoning performance would embody aspects of a calculus of probability, such as Bayes' Theorem. Support for this prediction was provided by the correct solutions reported by the majority of subjects receiving the frequency presentation. However, the accuracy rates generated in the current study were substantially lower than those reported by Cosmides and Tooby. Furthermore, explicit examination of subjects' descriptions of their solutions indicated the reasoning performance of the majority of subjects, regardless of presentation, embodied very little of Bayes' Theorem. As well, some subjects committed errors in reasoning that would not have been detectable in the Cosmides and Tooby study. Taken together, these results do not support the conclusion that humans are good intuitive statisticians.

As described by Gigerenzer, the model of the mind as an intuitive statistician arose after the incorporation of statistics as an indispensable part of the experimental method of psychology (Gigerenzer, 1994; Gigerenzer et al., 1989; Gigerenzer & Murray, 1987). It was only then that rational thought became defined by the laws of probability. The prevailing view of early research in human thought processes, as evidenced in the work of Piaget (1967), was that by adolescence, the mind was a reasonable intuitive statistician. However, empirical investigations of reasoning in the 1960s and 1970s, which incorporated Bayesian probability as the rational standard, painted a different picture. Studies with urn and ball problems required subjects to judge the probability that a sample was from an urn with a predominant colour of balls, given the colour of the balls drawn. Subjects were found to be conservative Bayesians, revising their probability estimates to a lesser extent than Bayes' theorem would dictate on the basis of the data they observed (Edwards, 1968). This line of research was fruitfully extended to real-life situations with the judgment under uncertainty program of Kahneman and Tversky (1973; Kahneman, Slovic, & Tversky, 1982). The substantial body of data generated suggested that, rather than conforming to the laws of probability, human reasoning was characterized by numerous fallacies and biases suggestive of the operation of heuristics.

Gigerenzer and Cosmides and Tooby have attacked both the theoretical and empirical foundation of this view of human reasoning. Their criticisms centre on the belief that the acceptance of Bayesian probability as the sole standard for the rational mind is inappropriate. Gigerenzer (1991b) has identified three errors in the judgment under uncertainty program; once these errors are corrected a different view of human reasoning emerges. First, he argues that some of these studies have presented one normative answer against which subjects' responses are compared, when in fact several *correct* statistical solutions are possible and defensible. Kahneman and Tversky commit this error with the Cab Problem. When subjects' responses are compared to these alternative, but equally valid, standards of correctness base rate neglect disappears (Gigerenzer & Murray, 1987). Second, assumptions that must hold true for the valid application of probability theory are neglected within some studies. One such example is research using the Engineer/Lawyer problem (Kahneman & Tversky, 1973), which supported base rate neglect. With some versions of this problem, subjects were not told that random sampling, a necessary condition for the base rates to be relevant, had occurred. In other versions, subjects were very briefly told that the descriptions used had been randomly sampled when in fact this was not the case. Gigerenzer, Hell, and Blank (1988) redid this experiment, making the subjects explicitly aware of the randomness of the descriptions by having them draw the descriptions blindly from an urn. Subsequently, base rate neglect disappeared.

Finally, and most importantly, Gigerenzer (1991; 1994) and Cosmides and Tooby (in press) argue that these studies have neglected the distinction between subjective Bayesian and frequentist schools of probability theory. This final error has two implications for the "heuristics and biases" findings. First, in light of this distinction, some of the reported biases, such as overconfidence, are no longer sensical. Overconfidence refers to the finding that subjects' confidence in their judgments are greater than the relative frequency of correct answers. However, Gigerenzer notes that this bias reflects a discrepancy between subjects' judgments of a single-event probability and a relative frequency, and that "even a subjectivist would not generally think of a discrepancy between confidence and relative frequency as a bias" (1991b, p. 261). Second, presenting the same problems with a frequentist rather than Bayesian framework has been shown to elicit improved reasoning. Subjects asked for relative frequency judgments instead of single-event probabilities do not evidence the overconfidence bias or the conjunction fallacy (Gigerenzer, 1994). Similar conclusions have been drawn for base rate neglect on the basis of Cosmides and Tooby's findings with the Medical Diagnosis problem. These studies have been interpreted as supporting the hypothesis that human mental mechanisms have evolved to operate with frequencies, the manner in which information was probably presented and acquired in the ancestral environment.

However, both Gigerenzer and Cosmides and Tooby appear to be changing more than just the information representation in their frequentist versions of the Medical Diagnosis problem. Gigerenzer (1994) presents the following example of a Bayesian version of the problem:

> The prevalence of breast cancer is 1% (in a specified population). The probability that a mammography is positive if a woman has breast cancer is 79%, and 9.6% if she does not. What is the probability that a woman who tests positive actually has breast cancer? _____% (p. 146)

He then suggests the following thought experiment:

> Change the information representation in the mammography problem from single-event probabilities to frequencies:
>
> Imagine 100 people (think of a 10 X 10 grid). We expect that one woman has cancer and a positive mammography. Also, we expect that there are 10 more women with positive mammographies but no cancer. Thus we expect 11 people with positive mammographies. How many women with positive mammographies will actually have breast cancer?
>
> With frequencies, you immediately "see" that only about *1 out of 11* women who test positive will have cancer. The base rate fallacy disappears if the information is represented in frequencies. (p. 146)

But it is not just the representation of the information that has changed between these two versions. In the frequency version, an explicit sample has been provided, as well as an instruction that encourages visual representation of that sample. The true positive rate has been deleted, and the application of an implicit 100% true positive rate to the base rate has already been incorporated. The numerator component of Bayes' theorem has been calculated for the reader. Similarly, the reader is also relieved from having to apply the false positive rate to the correct sample, as the problem provides the result of this application, the number of false positives in this sample. Misapplication of the false positive rate is one of the errors subjects in the current study were found to commit. In addition, these two components have been combined to provide the total number of positives in the sample, information which has no comparable counterpart in the Bayesian percent version. Thus, in addition to changing the information to frequency form, Gigerenzer has also completed all but the final step of the problem for the reader. Is it any wonder that one "immediately sees" the solution?

In the present investigation, when the effect of the frequency presentation in the Medical Diagnosis problem was isolated from the effects of confounding variables, subjects no longer appeared to be good intuitive frequentist statisticians. Thus, the conclusion that frequency presentations eliminate base rate neglect appears premature. It may well be that errors in the application of probability theory within the judgments under uncertainty program are responsible for some of the biases that have been found. However, whether our evolved reasoning mechanisms operate as frequentists is a separate issue.

The frequentist hypothesis of our inductive reasoning mechanisms is grounded in the findings that the human mind represents probabilistic information as frequencies for many domains (Cosmides & Tooby, in press). It seems obvious that frequencies are easier to comprehend than probabilities. But is this necessarily reflective of our mental mechanistic design, or simply the fact that the concept of frequencies involves a lower level of abstraction than that of probability? Cosmides and Tooby have suggested that only certain cognitive mechanisms might be expected to be frequentist in design, namely those that operate in domains where "event frequencies are observable, are relevant to the problem, and are the sole, the primary, or the best source of information available for solving the problem" (in press, manuscript p. 70). However, whether or not the Medical Diagnosis problem fits this description is debatable. It is quite probable that a disease prevalence rate of 1 out of 1000 would not be observable within the context of the ancestral hunter-gatherer lifestyle, where the number of individuals encountered in a lifetime was much smaller than that today. Therefore, it may be that the Medical Diagnosis problem is a poor choice for testing the frequentist reasoning hypothesis.

Finally, it would be useful to clarify how the frequency presentation is having an effect, in cases where an effect is observed. Is it by providing the information in a form with which our mental mechanisms can function, as suggested by Gigerenzer and Cosmides and Tooby, or by guiding our conception of the nature of the problem? The finding from Experiment 4 that subjects receiving the frequency presentation actually showed poorer understanding of the Bayes' Theorem solution than those receiving the percent presentation may be relevant to this issue. If the frequency presentation had originally facilitated correct responses by structuring the problem as the categorization of different individuals in the group of 1000, this effect could have backfired with the new rates. Presenting the information as a frequency may have actually confused subjects as to the distinction between these groups of 100, and facilitated an inappropriate categorization. Although this mistake would not produce an error with the original rates, the increased base rate in the problem in Experiment 4 meant that this error would now produce an incorrect result. This interpretation would be consistent with the finding of reduced reasoning. Because the percent presentation did not include the references to the group sizes, the same effect would not be expected.

Furthermore, some of Cosmides and Tooby's frequency versions of the Medical Diagnosis problem encouraged subjects to form a pictorial representation of the sample. These versions elicited the highest accuracy rates. Similarly, Gigerenzer's frequency version quoted above also suggests a visual representation. Although Cosmides and Tooby have attributed the resulting increase in accuracy to the concreteness of the frequency representation, it is possible that these visual representations aid in structuring the problem as one of categorization.

The conclusion that errors in reasoning disappear with frequency presentations was not supported by this research with the Medical Diagnosis problem. This contradicts the finding of a frequency presentation effect reported with other judgment under uncertainty problems. Further investigation appears necessary to determine when and how frequency presentations influence reasoning performance, and whether these results support the hypothesis that human reasoning is governed by frequentist mental mechanisms.

References

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

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

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

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), Formal representations of human judgment. New York: Wiley.

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

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

Gigerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. In G. Keren and C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues (pp. 311-339). Hillsdale, NJ: Erlbaum.

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

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

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

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

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

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

Kahneman, D. & Tversky, A. (1972). On prediction and judgment. ORI Research Monograph, 12.

Lunneborg, C. (1994). Modeling experimental and observational data. Belmont, CA: Duxbury Press.

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

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

Piaget, J. (1967). Six psychological studies. New York: Random House.

Appendix A - Sample Questionnaires

Experiment 1

Original Problem/Percent Information

PCB/E1C1

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?

______%

<u>Original Problem/Frequency Information</u>

FB/E1C2

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 1 out of 1000 has a false positive rate of 50 out of 1000, on average, how many people who test positive for the disease will actually have the disease, assuming you know nothing about their symptoms or signs?

______ out of ______

<u>Original Problem/Decimal Probability Information</u>

PBB/E1C3

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 0.001 has a false positive rate of 0.05, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?

______ (record your answer as a decimal probability)

Clarified Problem/Percent Information

PCC/E2C1

Please read the problem carefully before answering any questions.

1/1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, of all people who are perfectly healthy, 5% will test positive for the disease.

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:
on average,

What is the chance that a person found to have a positive result actually has the disease?

______%

Clarified Problem/Frequency Information

FC/E2C3

Please read the problem carefully before answering any questions.

1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, for every 1000 people who are perfectly healthy, 50 of them will test positive for the disease .

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of these people.

Given the information above:
on average,

How many people who test positive for the disease will actually have the disease?

______ out of ______

Clarified Problem/Decimal Probability Information

PBC/E2C2

Please read the problem carefully before answering any questions.

The rate of disease X among Americans is 0.001. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, for people who are perfectly healthy, the probability of testing positive for the disease is 0.05.

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:
on average,

What is the chance that a person found to have a positive result actually has the disease?

______ (record your answer as a decimal probability)

Experiment 2

Clarified Problem/Percent Information:

PCC/E2AC1

Please read the problem carefully before answering any questions.

1/1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, of all people who are perfectly healthy, 5% will test positive for the disease.

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:
on average,

What is the chance that a person found to have a positive result actually has the disease?

______%

Clarified Problem/Frequency Information

FC/E2AC3

Please read the problem carefully before answering any questions.

1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, for every 1000 people who are perfectly healthy, 50 of them will test positive for the disease .

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:
on average,

How many people who test positive for the disease will actually have the disease?

______ out of ______

<u>Cosmides & Tooby's Clarified Problem/Frequency Information</u>

FC/E2AC2

Please read the problem carefully before answering any questions.

1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease .

Imagine that we have assembled a random sample of 1000 Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:
on average,

How many people who test positive for the disease will *actually* have the disease?

______ out of ______

Experiment 3

Frequency Information (Base Rate 1/1000)/Common Denominator

FBCD/E1C2

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 1 out of 1000 has a false positive rate of 50 out of 1000, on average, how many people who test positive for the disease will actually have the disease, assuming you know nothing about their symptoms or signs?

______ out of ______

Frequency Information (Base Rate 1/100)/Common Denominator

FBCD/E3C1

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 1 out of 100 has a false positive rate of 5 out of 100, on average, how many people who test positive for the disease will actually have the disease, assuming you know nothing about their symptoms or signs?

______ out of ______

Frequency Information (Base Rate 1/1000)/No Common Denominator

FBNCD/E3C2

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 1 out of 1000 has a false positive rate of 5 out of 100, on average, how many people who test positive for the disease will actually have the disease, assuming you know nothing about their symptoms or signs?

______ out of ______

Percent Information (Base Rate 1/100)/Common Denominator

PCBCD/E3C3

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 1/100 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?

______%

Percent Information (Base Rate 1/1000)/No Common Denominator

PCBNCD/E1C1

Please read the problem carefully before answering any questions.

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has

______%

Experiment 4

Percent Presentation/100% True Positive Rate

PCEC/E4C1

Please read the problem carefully before answering any questions.

20/100 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, of all people who are perfectly healthy, 50% will test positive for the disease.

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:
on average,

What is the chance that a person found to have a positive result actually has the disease?

______%

<u>Frequency Presentation/100% True Positive Rate</u>

FEC/E4C2

Please read the problem carefully before answering any questions.

20 out of every 100 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, for every 100 people who are perfectly healthy, 50 of them will test positive for the disease .

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of these people.

Given the information above:
on average,

How many people who test positive for the disease will actually have the disease?

______ out of ______

Percent Presentation/80% True Positive Rate

PCECTP/E4C3

Please read the problem carefully before answering any questions.

20/100 Americans has disease X. A test has been developed to detect when a person has disease X. When the test is given to a person who has the disease, the test comes out positive 80% of the time. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, of all people who are perfectly healthy, 30% will test positive for the disease.

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:
on average,

What is the chance that a person found to have a positive result actually has the disease?

______%

Frequency Presentation/80% True Positive Rate

FECTP/E4C4

Please read the problem carefully before answering any questions.

20 out of every 100 Americans has disease X. A test has been developed to detect when a person has disease X. When the test is given to a person who has the disease, the test comes out positive 80 times out of 100. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, for every 100 people who are perfectly healthy, 30 of them will test positive for the disease .

Imagine that we have assembled a random sample of Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of these people.

Given the information above:
on average,

How many people who test positive for the disease will actually have the disease?

______ out of ______

Questionnaire Completed by All Subjects

Please describe how you arrived at your answer to the problem. Show any steps or calculations you used in the process. Please answer as completely as possible. **(Please do not go back and change your answer on the previous page in any way)**

The following questions relate to how you solved the problem. Feel free to look back at your answer if necessary. However, **it is important that you do not change your answer on the previous page.**

1. The following is a list of several possible ways to analyze the information presented in the problem. Please **circle the letter** beside any of the following combinations that you think are necessary for solving the problem. **Only circle the letter beside those combinations that must be considered to solve the problem.** Do not mark combinations that you do not need to consider.

A	the rate of diseased individuals in the population	multiplied by	the rate of testing positive when healthy
B	the rate of diseased individuals in the population	multiplied by	the rate of testing positive when diseased
C	the rate of testing positive when diseased	multiplied by	the rate of testing positive when healthy
D	the rate of healthy individuals in the population	multiplied by	the rate of testing positive when healthy
E	the rate of healthy individuals in the population	multiplied by	the rate of testing positive when diseased
F	the number of individuals in the population	minus	the number of healthy individuals who test positive
G	the number of individuals in the population	minus	the number of diseased individuals who test positive
H	the rate of testing positive when healthy	minus	the rate of diseased individuals in the population

2. Did you consider any combinations that were **not included** in the previous list? YES NO

If yes, list the other combinations you considered:

3. What did you do with the combinations you indicated above to get your answer? Please be as specific as possible.

4. Have you taken any courses in the following areas? Please write the number of courses you have taken in each area in the space given.

_______ Archaeology
_______ Biology
_______ Business or Economics
_______ Chemistry
_______ History
_______ Languages
_______ Mathematics
_______ Physics
_______ Psychology
_______ Statistics

5. What is your cumulative GPA (CGPA)? __________

Appendix B - Sample Coding Protocol for Raters

Coding Guide:

The following calculation illustrates the solution to the problem using Bayes' theorem:

$$P(D/+) = \frac{P(D) * P(+/D)}{[P(D) * P(+/D)] + [P(\sim D) * P(+/\sim D)]}$$

where:

P(D/+) = the probability of having the disease if you have tested positive
P(D) = the probability of having the disease (the rate of the disease in population)
P(+/D) = the probability of testing positive if you have the disease (the true positive rate)
P(~D) = the probability of not having the disease
P(+/~D) = the probability of testing positive if you do not have the disease (the false positive rate)

Solving the problem correctly requires the subject to:

a) Calculate the number of individuals who would **have the disease and test positive** (this is the numerator above)
b) Calculate the number of individuals who would **not have the disease and test positive.**
c) Add these two numbers to get the **total number of individuals testing positive** (this is the denominator above).
d) Divide the numerator by the denominator to get the number who would **have the disease and test positive out of the total number who would test positive.** This provides the chance of having the disease if you test positive.

In Conditions 1 & 2 the answer would look like this:

$$P(D/+) = \frac{20 * 100\%}{(20 * 100\%) + (80 * 50\%)} \quad \text{OR} \quad \frac{0.20 * 1.00}{(0.20 * 1.00) + (0.80 * 0.50)}$$

$$= \frac{20}{20 + 40} \qquad \frac{0.20}{0.2 + 0.4}$$

$$= \frac{20}{60} \qquad \frac{0.2}{0.6}$$

(Note that subjects may not bother to include the 100% true positive rate, as it does not affect the numbers)

In Conditions 3 & 4 the answer would look like this:

$$P(D/+) = \frac{20 * 80\%}{(20 * 80\%) + (80 * 30\%)} \quad \text{OR} \quad \frac{0.20 * 0.80}{(0.20 * 0.80) + (0.80 * 0.30)}$$

$$= \frac{16}{16 + 24} \qquad \frac{0.16}{0.16 + 0.24}$$

$$= \frac{16}{40} \qquad \frac{0.16}{0.40}$$

Definitions for coding the answer:

<u>For C1 & C2</u> - Correct answer = 33.3%
0.33
20 out of 60
1 out of 3
or equivalents

<u>For C3 & C4</u> - Correct answer = 40%
0.40
16 out of 40
40 out of 100
or equivalents

*****However, it is possible to come up with an answer of 40% without solving the problem correctly. For example, simply computing the total number of positives in the sample would also result in an answer of 40 out of 100. Therefore, it is important to verify that the 40% answer reflects the correct solution, and not an error in reasoning. **If it reflects an error, code the answer as incorrect.**

<u>For all conditions</u>:

Complete Bayes' Theorem reasoning	=	Calculates numerator and denominator components, and combines these appropriately
Partial Bayes' Theorem reasoning	=	Calculates the numerator and denominator components, but combines them incorrectly, OR applies a correct Bayes' Theorem framework but miscalculates component parts
No Bayes' Theorem reasoning	=	Explanation of answer shows very little to no evidence of Bayes' Theorem components or framework.

Please record the following information on the Coding Chart for each questionnaire:

Page 1:

A1 - Record the subject's reported answer to the problem

Page 2:

A2 - Record the subject's answer to the problem given in the explanation

Answer (**AN**)- Code the subject's answer as either:
1 = correct
2 = incorrect

Reasoning (**RE**) - From the explanation of the answer, code the subject's reasoning as one of the following:
1 = complete Bayes' Theorem reasoning
2 = partial Bayes' Theorem reasoning
3 = no Bayes' Theorem reasoning

Question Being Answered (**Q2**)- Code the subject's interpretation of the question being asked as one of the following:

1 = (the # who have disease and test positive) out of (the total # testing positive)
2 = (the # who have the disease & test positive) out of (the # who have disease)
3 = (the total # testing positive) out of (the total # in the sample)
4 = (the # who have the disease) out of (the total # in the sample)
5 = (the number who test positive) out of (the number who have the disease)
6 = unclear
7 = (the # who do not have the disease and test positive) out of (the total # in the sample)

Q1 - If the subject has changed their answer from A1 to A2, it is possible that they interpreted the question differently at A1. If possible, code the interpretation of the question at A1 on the scale above, and record this as Q1

Page 3:

Abstract understanding of Bayes components (**AB**)- Code the subject's identification of necessary combinations on the following scale:
1 = B & D
2 = B & D & others
3 = B or D (but not both)
4 = B or D (but not both) & others
5 = neither B nor D

Page 4 - Question 3:

A3 - If the subject has changed their answer to the problem, report the subject's latest answer to the problem as A3. If no information is given for Question 3, or if the information given does not include a new answer to the problem, report A3 as the same as A2.

Manipulation (**MA**) - Record what the subject reports doing with the combinations identified as either:

1 = correctly manipulated components according to Bayes theorem
2 = did not manipulate components according to Bayes theorem
3 = manipulations show some Bayes reasoning