Memory & Cognition
2006, 34 (3), 603-618

Use of base rates and case cue information in making likelihood estimates

STEPHANIE STOLARZ-FANTINO, EDMUND FANTINO, and NICHOLAS VAN BORST
University of California, San Diego, La Jolla, California

In five experiments, we investigated college students' use of base rate and case cue information in estimating likelihood. The participants reported that case cues were more important than base rates, except when the case cues were totally uninformative, and made more use of base rate information when the base rates were varied within subjects, rather than between subjects. Estimates were more Bayesian when base rate and case cue information was congruent, rather than contradictory. The nature of the "witness" in case cue information (animate or inanimate) did not affect the use of base rate and case cue information. Multiple trials with feedback led to more accurate estimates; however, this effect was not lasting. The results suggest that when base rate information is made salient by experience (multiple trials and within-subjects variation) or by other manipulations, base rate neglect is minimized.

Base rate neglect refers to the robust finding that people often underweight the importance of base rates in a decision task involving two or more sources of information (see, e.g., Goodie & Fantino, 1996; Koehler, 1996; Tversky & Kahneman, 1982). In base rate experiments, participants are typically provided with base rate information, which concerns how often each of two outcomes occurs in the general population, and case-specific information, such as the results of a diagnostic test or witness testimony. The participants' task, typically, is to select the more likely of two outcomes or to provide a verbal estimate of the probability of one or both outcomes. The *taxicab problem*, described by Tversky and Kahneman, is one of the most recognizable examples:

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (a) 85% of the cabs in the city are Green and 15% are Blue.
>
> (b) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.
>
> What is the probability that the cab involved in the accident was Blue rather than Green?

If both pieces of information were considered and were combined according to Bayes's theorem, the probability estimate would be close to .41; however, in most studies, adult participants overweight the case-specific information (b) and neglect the base rate information (a). This has been found when the probabilistic information has been conveyed verbally (as in Tversky & Kahneman, 1982) and also when it has been directly experienced (as, e.g., in the behavioral studies of Goodie & Fantino, 1996).

Underutilization of base rate information is of interest in that its occurrence bears on the general question of whether normative models of inference are also descriptive of human reasoning (see, e.g., M. S. Cohen, 1993; Doherty, 2003; Stanovich & West, 2000) and the more specific question of the extent to which people understand principles of probability and apply them in making judgments and decisions (e.g., Gigerenzer, 1998; Hertwig & Gigerenzer, 1999; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). More particularly, there is discussion and some disagreement as to when it is normative to use base rates as a factor in making likelihood judgments and decisions (e.g., Bar-Hillel, 1990; Birnbaum, 1983; L. J. Cohen, 1979).

Does it matter whether or not people make use of base rates when they judge likelihood? The use of base rates is of practical importance in, for example, the area of medical diagnosis. In a study in which physicians were asked to estimate the probability of a woman's having breast cancer, given a positive mammogram, Eddy (1982) reported that most estimated the probability at around 75%—close to the sensitivity of the test, which had been reported to the participants as 79%, and far from the correct answer of about 8%. Similar findings were reported by Gigerenzer, Hoffrage, and Ebert (1998), who arranged for visits

This research was supported by NIMH Grant MH 57127 to the University of California, San Diego. The authors thank the following students for their assistance: Kelloir Smith, Karmen Madatyan, Edward Hsu, and Tammie Hsieh. Correspondence concerning this article should be addressed to E. Fantino, Department of Psychology, 0109, University of California, San Diego, La Jolla, CA 92093-0109 (e-mail: efantino@ucsd.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

to 20 German public health clinics, where counselors were asked about the likelihood that a man who had no risk factors but who tested positive for HIV was actually infected with the virus. They found that most counselors asserted that false positives did not occur; usually, they, like Eddy's participants, confused the likelihood of infection with the sensitivity of the test. Do such findings mean that physicians routinely ignore base rates when judging the relative likelihood of two diagnoses? Not necessarily, especially when they have relevant past experience (Christensen-Szalanski & Bushyhead, 1981; Weber, Böckenholt, Hilton, & Wallace, 1993). However, Kennedy, Willis, and Faust (1997) reported that although school psychologists used base rate information appropriately so long as no other clinical information was available, their use of base rates—and consequently, their diagnostic accuracy—fell when other diagnostic information was added, whether the information was relevant or irrelevant. Base rate information can be important not only when clinical judgments are made, but also in other situations—for example, in legal settings, where jurors are asked to weigh evidence under conditions of uncertainty (see, e.g., Kaye & Koehler, 1991, 2003).

Although much is known about the determinants of base rate neglect (see, e.g., Koehler, 1996), interesting questions remain. Some of these questions concern how participants' likelihood estimates are affected by the following factors.

1. *The format in which problems are presented*. Much attention has been given to this issue. Most notably, Gigerenzer and his associates (e.g., Gigerenzer, 1998; Gigerenzer & Hoffrage, 1995) have demonstrated that base rate problem performance improves markedly when the problems are presented in a natural frequency format, rather than in the more commonly used probability format. It would appear that the natural frequency format makes directly available to participants the information they need to make likelihood estimates, rather than forcing them to derive the information from base rates and rates of witness accuracy. Similarly Macchi (1995, 2000) has argued that making transparent the conditional probabilities inherent in base rate problems and the links among the pieces of information presented in problems may minimize base rate neglect. Certainly, it is desirable to present information in the most transparent and usable manner possible; however, much of the statistical information to which people are exposed in daily life is in the form of probabilities. They use this information to judge the likelihood of many events, ranging from the mundane (e.g., the weather) to the vitally important (e.g., the desirability of undergoing a medical procedure). Thus, it remains important to know more about how people make judgments when information is presented in a probability format; that format was used in the present experiments.

2. *Participants' experiences prior to presentation of the base rate problems*. It has been shown that participants' prior experience influences their performance on base rate problems. Thus, prior experience with the case-specific information may predispose people toward base rate neglect (e.g., Goodie & Fantino, 1996), whereas experience with the relevant base rates predisposes them to more accurate decisions (e.g., Case, Fantino, & Goodie, 1999). The results from the latter study also suggest that when attention is called to the causal nature of the base rates, participants are able to incorporate base rate information even when it is embedded in a less transparent manner. In any event, it is possible that calling attention to either base rate or case cue information might make that information more likely to be considered.

3. *How participants' responses are received*. Do participants receive feedback or reward for accurate responses? In a series of studies by Goodie and Fantino (1995, 1996, 1999a, 1999b), base rate neglect was assessed in an experiential (behavioral) task. They found that base rate neglect persisted for hundreds of trials, even with appropriate feedback for correct and incorrect responses and with reward (including monetary reward) for correct responses. However, the kind of feedback employed in these studies was not very directive or explicit. In some of the present experiments, we evaluated the effect of giving feedback about the quality of the participants' responses on base rate problem performance.

4. *The methodological context in which the issues are evaluated*. An important methodological point surfaced when Leon and Anderson (1974), using a within-subjects design, found that the judgments of participants in a two-urn task did not reflect base rate neglect; instead, judgments varied appropriately with changes in urn composition (corresponding to base rate) and sample size (corresponding to case cue accuracy). Similar findings have been reported by Ward (1975). This issue has also been discussed by Birnbaum and Mellers (1983), who showed that participants performed better at judging the probability of an event—that is, their judgments were closer to the normative response—when they were studied in a within-subjects, rather than a between-subjects, design. This topic will be considered further in the present work.

In the series of experiments reported here, we examined several issues. One was the effect of multiple trials with and without feedback on the accuracy of likelihood estimates. Goodie and Fantino found that in their behavioral task, several hundred trials were required before participants' performance improved to the point of minimizing base rate neglect (see, e.g., Goodie & Fantino, 1999b). In two of the present experiments, we examined repeated trials in which the information was conveyed verbally. The results pertain to the extent to which performance on base rate problems improves with repeated trials, with and without feedback about the correctness of the answers. We also examined whether within-session gains would be maintained in a follow-up test. Another issue we examined was whether increasing the salience of base rates, either by varying them within subjects or by making them more extreme, would lead participants to make more use of base rate information. Indeed, many of the variables that have been shown to modulate base rate neglect, including those introduced above, may do so by affecting the salience (or *transparency*) of the base rate information. In a similar vein, when witness information is patently unreliable—that is, when witness accuracy is described as 50% (in a two-choice task)—base rates should be particularly salient. Under these circumstances, would participants ignore the information given by the witness and rely on base rates? Our experiments also addressed the following questions. When witness accuracy is less than 50%, will participants adjust likelihood estimates in the direction opposite to the witness' opinion? Will participants' estimates be affected by whether the witness is described as a person or as a test or procedure? Will some participants make estimates that match those calculated according to Bayes's theorem?

In particular, in these experiments, we further explored the view that it is the salience of the base rates and the nature of prior experience that determine the extent to which base rates are neglected.

EXPERIMENT 1

In this experiment, we examined the effect on likelihood estimates of presenting repeated trials with varying base rates and case cue accuracies; the effect of feedback was also tested. The participants were given questions in a likelihood format and were asked to estimate the likelihood of a specified option with a number from 0 to 100. They received 24 questions, with varying base rates (20%, 40%, 60%, and 80%) and case cue witness accuracies (30%, 50%, and 70%); each of the 12 base-rate/witness-accuracy combinations was presented twice. Half of the participants received feedback after each of their answers, whereas the other half never received feedback. Would participants' accuracy improve over the 24 trials, and would feedback make a difference?

Method

Participants. Fifty-two students at the University of California, San Diego (UCSD), who were enrolled in psychology classes, served as participants, either to fulfill course requirements or to earn extra credit. There were 15 male and 37 female participants, ranging in age from 18 to 25 years; their mean age was 20 years, 2 months.

Materials. Twenty-four questions were used in the experiment. They involved different scenarios, but each had the following format:

> A deer was injured in an accident in a state park. A witness reported the accident to the park ranger. You are given the following data:
>
> A. 40% of the deer in the area are black-tailed deer and 60% are white-tailed deer.
>
> B. The witness identified the deer as a black-tailed deer. In the past, this person has correctly identified each of the two types of deer 30% of the time and has incorrectly identified them 70% of the time.
>
> Estimate the likelihood that the deer really was a black-tailed deer by entering a number between 0 and 100, where "0" means that it definitely was not a black-tailed deer and "100" means that it definitely was a black-tailed deer.
>
> What is the likelihood that the deer was a black-tailed deer?

Procedure. The students participated individually in small experimental rooms, each equipped with a computer. The questions were presented in one of three randomly selected orders. The students were assigned alternately to a feedback or a no-feedback condition, and each was given a sheet of paper containing the following instructions:

> In this study, you will be presented with a series of 24 different situations, one at a time. For each situation, you will make a judgment about how likely it is that a particular statement is accurate. You will indicate your answer by typing in a number between 0 (definitely not accurate) and 100 (definitely accurate). If the computer assigns you to the "feedback" condition, you will get information about whether each response is in the range considered correct; otherwise, you will not get feedback about the correctness of each response. However, you may ask the experimenter to tell you afterwards how many of your responses were in the correct range.
>
> If you have any questions about the procedure, please ask. If you want to write anything down as you work, it is okay to use this piece of paper. Finally, if you have participated in a similar experiment before (using the computer), please tell the experimenter about it. Thank you for your participation.

The students answered each question by typing a number from 0 to 100 and pressing "enter." Those in the feedback condition received a feedback message and then pressed "enter" once again to initiate the following trial. Those in the no-feedback condition went directly to the next trial after entering their responses. There were two questions for each combination of base rate (20%, 40%, 60%, and 80%) and witness accuracy (30%, 50%, and 70%). No students in this or the subsequent experiments indicated that they had participated in a similar experiment before.

Results and Discussion

An ANOVA was performed to determine whether the participants' estimates of likelihood were affected by base rate and witness accuracy. The results showed that both factors affected estimated likelihood [for base rate, $F(3,153) = 110.31$, $MS_e = 141.00$; for witness accuracy, $F(2,102) = 102.22$, $MS_e = 230.79$; for the base rate × witness accuracy interaction, $F(6,306) = 7.91$, $MS_e = 93.72$; all $ps < .01$]. The amount of variance accounted for by the variables, measured by ω^2, was 22% for both base rate and for witness accuracy and 1.8% for their interaction. The median estimates are shown in Figure 1, along with the answers calculated using Bayes's theorem. It can be seen that the estimates increase in an orderly manner with increases in base rate and witness accuracy; the estimates tend to be greater than the Bayesian estimates when the base rate is low and smaller than the Bayesian estimates when it is high.

Answers were scored as *correct* if they were within 10 points on either side of the estimates determined by Bayes's theorem; the participants in the feedback condition were given feedback according to this criterion. The students answered two questions at each base-rate/witness-accuracy combination; thus, mean number correct could range from 0 to 24.

The mean number of questions correct overall was 11.92 ($SD = 4.14$), or just under 50%. The participants in the no-feedback condition answered a mean of 11.42 questions correctly ($SD = 4.53$), and those in the feedback condition answered 12.42 correctly ($SD = 3.73$); this is not a significant difference. However, it is more informative to check for improvement due to feedback by comparing the number correct on the first five questions with the number correct on the last five questions. Sign tests comparing scores on the first and last five questions in each group showed that the students not given feedback improved in 8 cases, did less well in 9 cases, and scored the

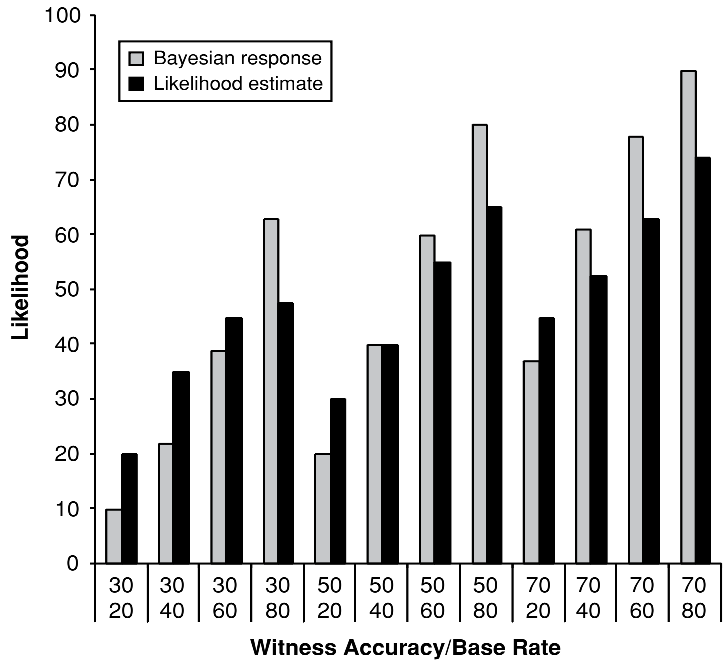

Figure 1. Experiment 1: Median likelihood estimates and corresponding Bayesian estimates.

same in 9 cases ($p > .99$). On the other hand, the students who were given feedback improved in 14 cases, did less well in 4 cases, and scored the same in 8 cases ($p < .03$). Thus, feedback seems to have had a positive but small effect on performance. Because feedback had essentially no effect on overall number correct and did not interact in a meaningful way with base rate and witness accuracy, it was omitted as a variable in the subsequent analysis.

Which base-rate/witness-accuracy combinations were easier or more difficult for the students to judge accurately? An analysis was performed to explore amount of error—defined as the absolute value of the difference between the optimal answer and the participant's answer—on each question as a function of base rate and witness accuracy. Means and standard errors are shown in the top part of Table 1. A 4 × 3 repeated measures ANOVA showed significant effects of base rate [$F(3,153) = 14.6$, $MS_e = 119.28$] and witness accuracy [$F(2,102) = 12.61$, $MS_e = 136.06$] and a significant base rate × witness accuracy interaction [$F(6,306) = 3.87$, $MS_e = 86.70$; all ps $< .01$]. The overall mean amount of error was 15.66 on the scale of 0 to 100 ($SE = 0.48$). The amount of error on 50% witness accuracy (WA 50) questions was significantly less than that on WA 30 and WA 70 questions, according to the Student–Newman–Keuls test ($\alpha = .05$). Since participants can, and should, ignore the unreliable wit-

Table 1
Differences Between Judged Likelihoods and Bayesian Likelihoods in Experiment 1

	Witness Accuracy							
	30		50		70			
Base Rate	*M*	*SE*	*M*	*SE*	*M*	*SE*	Mean	*SE*
			Absolute Values of Differences					
20	15.63	2.93	11.77	1.96	25.16	2.71	16.83	0.99
40	10.27	1.79	9.94	1.77	16.14	3.14	12.39	0.72
60	15.25	2.92	7.50	2.27	15.29	2.66	13.54	0.79
80	25.86	3.07	19.64	4.49	23.57	3.85	19.78	1.13
Mean	16.67	0.83	12.39	0.82	17.85	0.77	15.66	0.48
			Directional Differences					
20	14.14	2.22	10.31	1.84	6.98	2.65	10.47	1.32
40	12.06	1.57	1.64	1.39	−10.12	1.62	1.19	1.14
60	1.28	1.91	−7.77	1.66	−15.94	1.59	−7.48	1.14
80	−16.25	2.50	−16.96	2.23	−18.46	1.82	−17.22	1.26
Mean	2.81	1.33	−3.20	1.14	−9.39	1.20	−3.26	0.73

ness accuracy in the 50% questions, this result is to be expected. There were also significant differences among all levels of base rate, except for the comparison between the 40% base rate questions (BR 40) and BR 60 questions, both of which had lower levels of error than did the other questions. The participants were most accurate in their ratings of the BR 60/WA 30 and BR 40/WA 50 combinations. They were least accurate on all three of the BR 80 combinations, for which likelihoods were substantially underestimated. The bottom part of Table 1 shows directional error scores for all base-rate/witness-accuracy combinations. As can be seen, the participants overestimated likelihoods for BRs 20 and 40 and underestimated likelihoods for BRs 60 and 80.

Did any likelihood judgments correspond to the exact answer as predicted by Bayes's theorem? Answers were scored as *Bayesian* if they fell at or within 1 point of the likelihood calculated using Bayes's theorem. However, because the participants tended to make estimates that were rounded to the nearest 5, when the Bayesian answer was in between, the criterion was loosened. Thus, for example, if the Bayesian answer was 37, likelihood estimates of 35 and 40 were considered Bayesian. The participants' total numbers of Bayesian answers (out of 24 possible) ranged from 0 to 16, with a mean of 4.96 ($SE = 0.575$), a median of 4.0, and a mode of 1. The mean number of Bayesian answers in the feedback condition was 5.54 ($SE = 0.77$), and in the no-feedback condition, it was 4.385 ($SE = 0.86$); this was not a significant difference [$t(50) = 1.0, p > .1$]. Most of the participants gave few or no exact Bayesian answers; however, 5 of the 52 participants (9.6%) gave Bayesian answers at least half the time. An interesting finding from this within-subjects study is that base rate and witness accuracy were equally influential in contributing to likelihood estimates. That is, each factor contributed the same amount (22%) to the variance accounted for. Thus, these results do not single out base rate neglect (as compared with *witness neglect*) as the predominant source of error.

EXPERIMENT 2

Will giving students explicit training with repeated trials improve their performance on typical base rate questions, in the sense of making their likelihood estimates more Bayesian? And will such training cause them to develop a sense of how base rate and witness accuracy cues combine to determine likelihood? In Experiment 2, we explored these issues by giving the students a more prolonged exposure to base rate problems that included more explicit feedback than that used in Experiment 1 and by testing retention and generalization to new base rate questions.

Method

Participants. The participants were 27 students taking a summer session course in learning at UCSD. They volunteered for this project as one among several ways of fulfilling a course requirement to participate in and/or write about experimental research in learning. An additional 21 students who did not participate in the training served as a comparison group by answering the same follow-up questions as those given to the trained participants.

Materials. The questions used had the same format as those in Experiment 1 and, as before, were presented by computer. Only two witness cue accuracies were used in training: 30% and 70%. At each of these levels of witness accuracy, four levels of base rates were represented: 20%, 40%, 60%, and 80%. Additional *novel* (untrained) combinations of base rate and witness accuracy were used in testing, as will be described below.

Procedure. The students participated in three sessions over a period of 10 days. At the beginning of the first session, they were given the following instructions:

> In this study, during each session you will be presented with a series of different scenarios, one at a time. For each scenario, you will make a judgment about how likely it is that a particular statement is true. You will indicate your answer by typing in a whole number between 0 (definitely not true) and 100 (definitely true). After you make each judgment, you will get information about whether your response is in the range considered correct. If you are within 5 points on either side of the optimal answer, the computer will tell you that your answer is within the correct range. If you are within 10 points on either side of the optimal answer, the computer will tell you that your answer is slightly above or below the correct range. If you are more than 10 points off in either direction, it will tell you that your answer is significantly higher or lower than is correct.
>
> If you have any questions about the procedure, please ask the experimenter. If you want to write anything down as you work, you may use this piece of paper. Finally, if you have participated in a similar experiment before (using the computer), please tell the experimenter about it. Thank you for your participation.

Half of the participants were assigned to be trained on the WA 30 questions during Session 1 and the WA 70 questions during Session 2; the others received their training in the opposite order. During each of the two training sessions, the participants responded to 5 questions at BR 20, then 5 at BR 40, 5 at BR 60, and 5 at BR 80. They received feedback on their likelihood estimates, as described in the instructions. Then the sequence was repeated, once again working through BRs 20–80, always with the same witness accuracy, for a total of 40 base rate questions answered with feedback. Finally, at the end of each training session, the participants were tested with 4 questions from the base rates they had been trained on, plus 3 novel questions involving the same witness accuracy but base rates on which they had not been trained. For example, a student who had been trained on WA 30 and BRs 20, 40, 60, and 80 received the following novel questions: WA 30/BR 30, WA 30/BR 50, and WA 30/BR 70.

Session 3 was a testing-only session. The participants answered 20 questions, without feedback. Eight of these involved the witness-accuracy/base-rate combinations used in training. The other 12 questions were novel; of these, 6 had not been presented in training but had been presented in testing (WAs 30 and 70 combined with BRs 30, 50, and 70). An additional 2 questions represented BRs 30 and 70 combined with WA 50. Finally, 4 questions involved more extreme base rates (BRs 10 and 90) combined with WAs 30 and 70.

After completing the test questions, the participants answered two follow-up questions. One question was a likelihood-format scenario similar to those on which the participants had been trained and tested previously:

> While camping in a state park in the Sierras, a visitor finds a gold-colored nugget in a stream. He shows it to a park ranger. You are given the following two pieces of information:
>
> In the past, 10% of such nuggets have been found to be gold and 90% to be iron pyrite.
>
> The ranger identifies the nugget as gold. In the past, he has correctly identified gold and iron pyrite 60% of the time and has incorrectly identified them 40% of the time.

> _______ Estimate the likelihood that the nugget was gold by entering a number from 0 to 100, where 0 means that the nugget definitely was not gold and 100 means that it definitely was gold.

The second question was in natural frequency format, as follows:

> Professor Snipe has suggested that all students majoring in psychology should have to pass his "fear of math" screening test before being admitted to the department's statistics course. He says that students who score positively on the "fear of math" test will fail the statistics course. Screening them out would save a great deal of trauma. You are given the following information:
>
> Out of 100 psychology majors, 10 usually fail statistics on the first try.
>
> Out of the 10 who fail, 9 test positive for "fear of math" on Dr. Snipe's screening. However, out of the 990 students who pass statistics, 9 will test positive for "fear of math."
>
> Estimate the accuracy of Dr. Snipe's test as follows:
>
> ______% of students who test positive for "fear of math" will actually end up failing statistics.

The numbers provided corresponded to a base rate of 10% and a witness accuracy of 90%.

After completing each question, the participants were asked to write a short description that would "explain the reasoning behind your answer as though you were explaining it to another student."

Results and Discussion

On both of the first 2 days of the study, there was improvement over the course of the training session. On each day, the participants scored significantly higher—in terms of the number of correct answers, as defined in the instructions—on the second round of training questions than on the first [Day 1, $t(26) = 3.5$; Day 2, $t(26) = 2.95$; ps $< .01$, one-tailed]. There was also a slight, although not statistically significant, improvement in the students' test scores over the 2 training days: The average test score on Day 1 was 52%, and on Day 2, 61% [$t(27) = 1.61$, $p = .059$]. There was no significant difference between those students who received the WA 30 questions first and those who received the WA 70 questions first. Overall, the mean test scores obtained by the participants on their WA 30 and WA 70 days were identical: 56.6% correct, including the novel questions on which the participants had not been trained. The symmetry of these findings shows that symmetrical base rates and witness accuracies led to comparable performances, in terms of percent correct.

The results on the test-only day (Day 3), which often took place several days after the training sessions, showed poorer performance, in comparison with the earlier sessions, with a mean overall score of only 28.5%. This was true for both the previously trained questions ($M = 28.7\%$) and the novel questions ($M = 28.4\%$). An ANOVA on type of question was statistically significant [$F(2,52) = 5.69$, $MS_e = 0.03$, $p < .01$]. The Student–Newman–Keuls test ($\alpha = .05$) confirmed that the participants were more likely to answer the extreme base rate questions correctly ($M = .426$) than either the previously trained questions ($M = .287$) or the untrained questions that did not include extreme base rates ($M = .284$).

Did students retain nothing from their training with varying base rates and witness accuracies? Answers to the two follow-up questions were scored for numerical accuracy according to Bayes's theorem and also for the quality of the explanations given. The results showed that the students were, in fact, not ignorant of the necessity of considering both base rate and witness accuracy in making their estimates. In their answers to the likelihood question, 20 of the 25 participants (80%) for whom follow-up data were available specifically referred to base rate and witness accuracy (by mentioning the corresponding numbers) in their explanations; of the 5 participants who did not consider both factors, 2 considered base rate only, 1 considered witness accuracy only, and 2 explained that their responses were specific answers recalled from training (and, in these cases, misapplied). According to the criterion of accuracy applied in Experiment 1 (10 points on either side of the Bayesian response), 11 of the 25 participants (44%) answered correctly; 6 (24%) were within 5 points of the Bayesian answer (the criterion for *correct* in this study). However, there was no significant relationship between mentioning base rate and witness accuracy in the explanation and giving a likelihood estimate close to the Bayesian answer. On the frequency format question, 13 of the 25 participants (52%) gave the exact Bayesian answer, which was simple to calculate from the information given. All but 1 of these participants gave clear and correct explanations of their reasoning on the task. Of the 12 participants who did not give the Bayesian answer, none gave explanations that would have led to a correct answer; thus, no answers were incorrect due to mistakes in calculation. There was no significant relationship between giving an estimate close to the Bayesian answer in Question 1 (likelihood format) and giving the Bayesian response to Question 2 (frequency format).

The same follow-up questions were also given to a group of 21 students who had not participated in Experiment 2. Among these participants, 14 (67%) gave answers within 10 points of the Bayesian answer on Question 1 (5 of these [24%] within 5 points), and 9 (43%) gave exact Bayesian answers to Question 2. Again, there was no significant relationship between the correctness of the students' answers to Question 1 and their answers to Question 2. Nor was there a significant difference between the proportions of experimental and comparison participants who answered each question correctly [$\chi^2(1, N = 46) = 2.36$ for Question 1 and 0.38 for Question 2; ps $> .1$]. As was the case with the experimental participants, the comparison participants who answered Question 2 correctly gave clear and correct explanations for their answers. However, on Question 1, the comparison participants were less likely to mention both base rate and witness accuracy as important for judging likelihood. Only 6 of the 21 comparison participants (29%) did so, in comparison with 80% of the experimental participants, a statistically significant difference [$t(40) = 4.006$, $p < .01$, two-tailed].

In summary, there is evidence that repeated exposure to changing base rates and to different levels of witness accuracy sensitized the participants to the importance of these factors. Despite this, there is no evidence that they developed an accurate understanding of the relationship of base rate and witness accuracy to likelihood, as calculated by Bayes's theorem, or that they performed more successfully than the participants without repeated exposure.

EXPERIMENT 3

Do people rely more or less on base rate or witness accuracy information when the "witness" is inanimate than when the witness is a person? It is conceivable that participants might consider an inanimate witness to be more reliable than a person and, therefore, might rely less on base rate information and more on the judgment of the witness when the witness is inanimate. For example, Jacobs and Potenza (1991) found that when given both base rate and individuating information, children and adults made more use of base rates in answering object domain questions than in answering social domain questions. Conversely, it could be argued that participants might consider a human witness more reliable and might be less likely to rely on base rates when the witness is a person. In the present experiment, a between-subjects design was used to examine the students' likelihood judgments of varied base-rate/witness-accuracy combinations; base rate salience was increased by including some base rates more extreme than those used in Experiments 1 and 2 (1% and 99%). In addition, in this experiment, we investigated whether the students would differ in their reliance on base rate or witness information depending on whether the "witness" was described as a person or was inanimate—that is, was described as a machine, procedure, or test.

Method

Participants. Five hundred sixty-six students taking undergraduate psychology courses at UCSD participated. As they entered their classrooms, each of the participants received one printed question, which they answered and turned in to the experimenters before the start of class.

Materials. Two types of questions were used. Those categorized *animate* had the same format as those used in Experiment 1; in each of the eight *animate* questions, the witness was described as a person. The eight questions making up the *inanimate* questions had the same format, except that the witness was a machine, procedure, or test. The levels of base rate given in the questions were 1%, 30%, 70%, and 99%; the levels of witness accuracy were 50% and 80%. All base-rate/witness-accuracy combinations were tested with both levels of question type (animate and inanimate). Each question utilized a different scenario; this was done in order to increase the plausibility of the base rates and witness accuracies included in the questions.

Procedure. The students were tested in their classrooms before their classes began. Participation was entirely voluntary. Each of the participating students received a half-sheet of paper with one question on it; there was enough space on the paper to carry out calculations if a student wished to do so. After reading the question, each participant wrote in a numerical likelihood estimate on a scale of 0 to 100 and returned the paper to the experimenters.

Results and Discussion

An ANOVA showed that both base rate and witness accuracy had a significant effect on the participants' estimates of likelihood [for base rate, $F(3,558) = 50.39$; for witness accuracy, $F(1,558) = 51.87$; $MS_e = 549.05$; both $ps < .01$]. There was no significant interaction of base rate and witness accuracy [$F(3,558) = 0.82$]. Variance accounted for, ω^2, was 19.4% for base rate and 6.6% for witness accuracy. Figure 2 shows the participants' median likelihood estimates, along with those calculated from Bayes's theorem.

In order to determine which base-rate/witness-accuracy combinations were easiest to judge, error scores were obtained by calculating the absolute value of the difference between the respondents' answers and the Bayesian answers; means and standard errors for the error scores of each base-rate/witness-accuracy combination appear in Table 2. The overall average deviation from the Bayesian answer was 27.32 points ($SE = 1.01$) on the 0 to 100 scale; there was no difference between the error scores for questions with inanimate witnesses ($M = 27.39$, $SE = 1.4$) and those for questions with animate witnesses ($M = 27.25$, $SE = 1.45$). A 4×2 ANOVA showed statistically significant effects of base rate [$F(3,558) = 29.50$, $p < .01$] and witness accuracy [$F(1,558) = 5.66$, $p < .05$], and a significant base rate $\times$ witness accuracy interaction [$F(3,558) = 9.71$, $MS_e = 472.65$, $p < .01$]. A Student–Newman–Keuls test ($\alpha = .05$), performed to evaluate differences in error between the levels of witness accuracy, showed that there was significantly more error on average for WA 80 questions than for WA 50 questions. There was also significantly more error in the answers to the BR 1 questions than to the BR 30, 70, or 99 questions. Except for a significantly larger amount of error on the BR 99 questions than on the BR 30 questions, the other base rates did not differ with respect to amount of deviation from the Bayesian response.

Finally, responses were scored according to whether they corresponded to the exact likelihoods calculated according to Bayes's theorem, using the criteria described in Experiment 1. Overall, 13.9% of the responses fit these criteria. There was a significant effect of base rate on the distribution of Bayesian answers. The greatest percentage of Bayesian answers, 25.9%, occurred for the BR 99 questions; next were BR 70, with 11.1%, and BR 1, with 11%. The fewest Bayesian answers, 6.5%, occurred for the BR 30 questions; the effect of base rate was statistically significant [$\chi^2(3, N = 596) = 33.56$, $p < .01$]. In agreement with the results of the error analysis, there were significantly more Bayesian responses on WA 50 questions (19.3% Bayesian responses) than on WA 80 questions (8.7% Bayesian responses) [$\chi^2(1, N = 566) = 13.1$, $p < .01$]. This result appears to demonstrate the participants' understanding that they should "go with the base rate" and ignore the witness when the witness was completely unreliable. It is surprising, however, that there were not an even greater number of Bayesian responses when witness accuracy was described as only 50% in a two-choice situation.

Whereas the BR 30 and BR 70 questions led to comparable performance (in terms of deviation from the Bayesian answer), the more extreme base rates (BR 1 and BR 99) led to very different performance: On BR 99 questions, 26% of the answers were Bayesian, as opposed to only

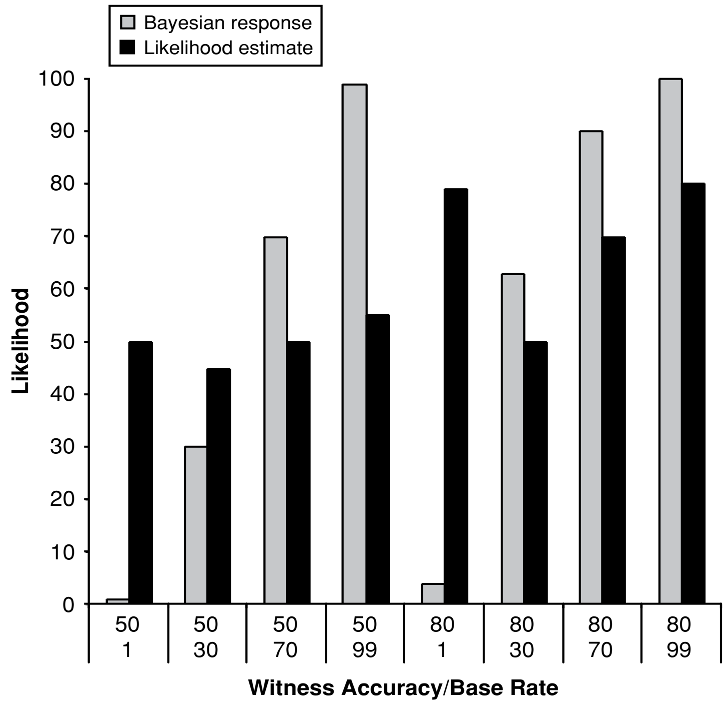

Figure 2. Experiment 3: Median likelihood estimates and corresponding Bayesian estimates.

11% for BR 1 questions. Also, the greatest degree of error occurred on the BR 1 questions. These results raise the possibility that when the base rate denotes an extremely likely event, it is less likely to be ignored than when it denotes an extremely unlikely event, although both base rates are equally informative. However, there is an alternative interpretation of this asymmetry: Since the witness accuracy of 80% was in much closer agreement with a base rate of 99% than with a base rate of 1%, superior performance might simply reflect this greater concordance of the two sources of information. The results from Experiment 4 should help to resolve this question.

EXPERIMENT 4

This experiment retested the base-rate/witness-accuracy combinations used in Experiment 1 but presented them in a between-subjects design. Thus, it should be possible to evaluate the effect of within-subjects variation of base rate—and also of witness accuracy—on the accuracy of the likelihood estimates. Two additional, more extreme base rates were also included, resulting in six levels of base rate (5%, 20%, 40%, 60%, 80%, and 95%) and three levels of witness accuracy (30%, 50%, and 70%). In this experiment, all the questions utilized the same scenario. It was hoped that this would reduce any variability that might be due to the use of particular scenarios, allowing a clearer evaluation of the effects of base rate and witness accuracy on likelihood estimates.

Method

Participants. The participants were 408 students: 145 men, 245 women, and 18 who did not identify themselves. All were enrolled in undergraduate psychology courses at UCSD.

Materials. The following question was used in this experiment:

> While walking in a canyon, a bird watcher saw a large black bird. She reported to the leader of her bird-watching group that she had seen a raven. You are given the following data:
>
> 1. 20% of the large black birds living in that canyon are ravens and 80% are crows.

Table 2
Differences Between Judged Likelihoods and Bayesian Likelihoods in Experiment 3

	Witness Accuracy					
	50		80			
Base Rate	*M*	*SE*	*M*	*SE*	Mean	*SE*
	Absolute Values of Differences					
1	33.57	2.92	50.33	3.69	42.36	2.47
30	26.22	1.06	25.14	2.61	19.11	1.17
70	19.76	1.78	22.38	1.46	21.29	1.13
99	30.90	3.28	20.66	2.91	22.45	2.21
Mean	24.90	1.26	29.69	1.56	27.32	1.01
	Directional Differences					
1	33.45	2.94	49.38	3.86	41.80	2.54
30	7.24	2.01	−13.37	4.49	0.56	2.19
70	−18.78	1.19	−22.26	1.48	−20.81	1.18
99	−30.81	3.29	−20.66	2.91	−25.41	2.22
Mean	−2.00	1.95	−1.71	2.35	−1.86	1.53

2. In the past, this bird watcher has correctly identified both ravens and crows 30% of the time and has misidentified them (that is, mixed them up) 70% of the time.

Estimate the likelihood that the bird was a raven by assigning a number from 0 to 100, where "0" means the bird definitely was not a raven (that is, it was really a crow) and "100" means the bird definitely was a raven.

All the participants received the same question, but the base rate and witness accuracy described in parts 1 and 2 were varied as described above.

Procedure. The procedure was the same as that in Experiment 3.

Results and Discussion

The median likelihood estimates for each combination of base rate and witness accuracy are shown in Figure 3, along with the answers calculated using Bayes's theorem. A 6 × 3 ANOVA performed on the estimates showed significant effects of base rate [$F(5,390) = 17.96$], and witness accuracy [$F(2,390) = 64.78$] and a significant base rate × witness accuracy interaction [$F(10,390) = 2.71$, $MS_e = 319.93$, all $ps < .01$]. Percentages of variance accounted for, measured by ω^2, were 13.3% for base rate, 20% for witness accuracy, and 3.2% for their interaction.

An ANOVA was performed on the error scores, where error was defined as the absolute value of the difference between the likelihood judgment and the answer calculated according to Bayes's theorem. The average amount of deviation (over all base-rate/witness-accuracy combinations) was 23.03 points ($SE = 0.93$) on the 0–100 scale. This is comparable to the average deviation of 27.32 found in Experiment 3. There was no difference between male and female participants in average amount of deviation from the Bayesian answer ($M = 23.6$ for males; $M = 22.9$ for females). There was a significant effect of base rate [$F(5,390) = 33.71$] and a significant base rate × witness accuracy interaction [$F(10,390) = 6.81$, $MS_e = 225.33$; both $ps < .01$]. However, there was no overall significant effect of witness accuracy [$F(2,390) = 2.125, p > .1$]; the mean error scores for WAs 30, 50, and 70 were 22.8, 20.6, and 25.7, respectively. Thus, the WA 50 questions did not yield more accurate estimates overall, although a specific comparison test showed the WA 50 and WA 70 questions to be significantly different from one another (Student–Newman–Keuls test, $\alpha = .05$). This result is in agreement with the finding in Experiment 3 of significantly more error in the estimates from WA 80 questions than in those from WA 50 questions. The questions containing BRs 5, 80, and 95 yielded significantly less accurate estimates than did those with other levels of base rate (Student–Newman–Keuls test, $\alpha = .05$). The least amount of deviation from the Bayesian answer was produced by the BR 40/WA 50 and BR 60/ WA 50 questions; those that produced the greatest amount of deviation were BR 95/WA 30, BR 5/ WA 70, and BR 5/WA 50. Error scores for all base-rate/ witness-accuracy combinations can be seen in Table 3. In general, when the base rates and witness cues were greatly divergent (e.g., BR 5/WA 70 and BR 95/WA 30), the deviation of judged likelihood from Bayesian likelihood was great. A corollary to this conclusion is that the participants' responses were less Bayesian when the witness cue and the base rates pointed to different conclusions—namely, when the two numbers were on different sides of 50%. This result helps us resolve the question of extreme but symmetrical base rates raised by the results of Experiment 3. We now

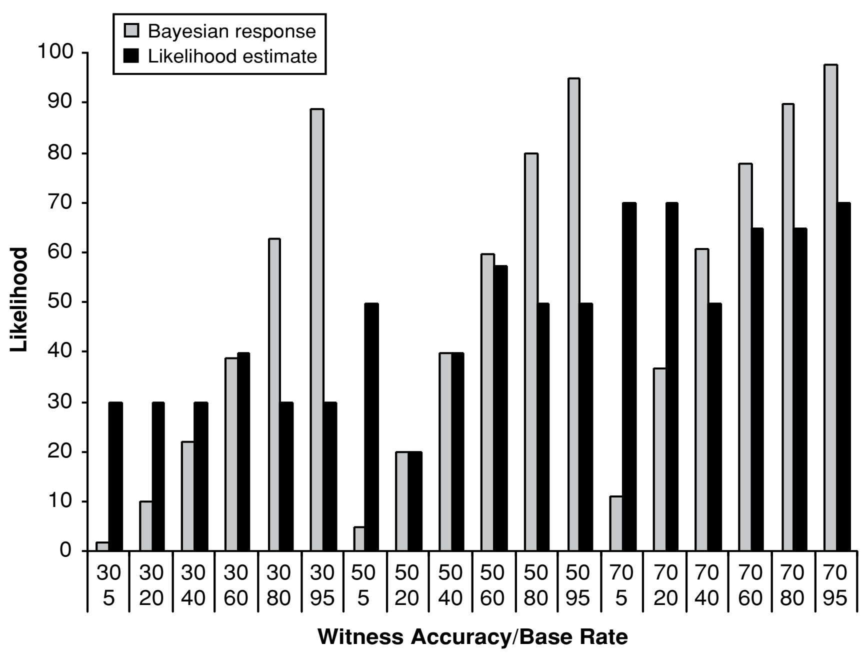

Figure 3. Experiment 4: Median likelihood estimates and corresponding Bayesian estimates.

Table 3
Differences Between Judged Likelihoods and Bayesian Likelihoods in Experiment 4

	Witness Accuracy							
	30		50		70			
Base Rate	*M*	*SE*	*M*	*SE*	*M*	*SE*	Mean	*SE*
	Absolute Values of Differences							
5	26.25	4.27	45.13	3.85	46.86	5.38	39.44	2.88
20	21.31	3.36	10.81	1.82	25.43	2.30	18.89	1.14
40	10.50	1.82	9.36	1.39	18.96	2.35	12.97	1.14
60	14.40	2.25	9.77	2.22	16.83	2.37	13.79	1.35
80	24.68	3.10	23.60	3.40	27.16	3.45	25.47	1.91
95	48.41	5.96	34.31	4.38	20.38	3.14	34.37	3.06
Mean	22.81	1.67	20.56	1.59	25.72	1.55	23.03	0.93
	Directional Differences							
5	26.15	4.30	45.13	3.85	44.95	6.14	38.74	3.04
20	20.00	3.67	4.55	2.62	17.69	4.38	13.79	2.18
40	7.33	2.12	−1.84	2.33	−13.04	3.66	−2.65	1.87
60	2.48	3.67	−5.68	2.81	−15.33	2.28	−6.07	2.00
80	−22.14	3.91	−23.60	3.40	−27.16	3.45	−24.39	2.05
95	−47.00	6.62	−34.31	4.38	−20.38	3.14	−33.90	3.16
Mean	−1.79	1.47	−0.37	2.59	−2.43	2.69	−1.79	1.47

see that when base rates and the witness cues are divergent, large errors occur, whether the extreme base rates are low (here, 5%) or high (here, 95%). When the base rate and the witness cue are less divergent (e.g., BR 95/WA 70 or BR 5/WA 30), much smaller deviations occur. There is no clear evidence of asymmetrical effects of extreme but symmetrical base rates.

Responses were also scored as to whether they fit the exact Bayesian answer. Overall, 12.25% of the estimates were so classified; this is comparable to the 13.9% found in Experiment 3. Did base rate influence how many Bayesian answers were given by the participants? The percentage of Bayesian answers by question base rate was as follows: BR 5, 5%; BR 20, 12.6%; BR 40, 17.6%; BR 60, 18.3%; BR 80, 11.1%; and BR 95, 5.7%. There was no significant tendency for base rates below 50% to have more or fewer Bayesian answers than did base rates above 50%. However, fewer Bayesian answers were given in answering the questions with extreme base rates (BR 5 and BR 95) than in answering those with more moderate base rates [$\chi^2(1, N = 408) = 6.47, p = .01$]. With respect to witness accuracy, there were fewer Bayesian answers when it was 30% (12.7% Bayesian answers) than when it was 50% (20.4%); there were especially few Bayesian answers for WA 70 questions (3.6%). These differences are statistically significant [$\chi^2(2, N = 408) = 17.99, p < .01$].

EXPERIMENT 5

In this experiment, we explored the question of how much importance the participants attributed to base rate and witness accuracy information. As in Experiments 3 and 4, each participant answered one base rate question; base rates and witness accuracies were varied across participants. The base rates tested were 5%, 40%, 60%, and 95%; the witness accuracies were 20%, 50%, and 80%. Each participant saw the same scenario, which involved deciding whether some berries picked by the protagonist were red or black elderberries. However, before making their likelihood estimates, the participants made judgments in which they divided 10 hypothetical points between two pieces of information: base rate and witness accuracy. In this way, they could express their opinions as to the relative importance of these pieces of information for indicating the likelihood that the berries were the variety identified by the protagonist.

Method

Participants. The participants were 362 students enrolled in undergraduate psychology courses at UCSD. They participated in their classrooms before the beginning of class activities.

Materials. The materials used the following scenario, with each participant receiving 1 of 12 possible combinations of base rate and witness accuracy:

> This questionnaire has two parts. Please read the information carefully and answer each part as accurately as you can. Thank you for your help.
>
> In the woods in Northern California, Martha picked some elderberries from a bush. Because only certain varieties are safe to eat, she needed to be sure what type they were. Martha identified the berries as black elderberries.
>
> **PART I**
>
> You are given **10 points** which you can assign to each of the following two pieces of data ("A" and "B"). On the line next to each, write the number of points (from 0 to ?) that you think reflects **the importance of that information** in deciding which type of elderberries Martha picked. For example, assigning no points to a piece of information indicates that it is not at all useful; assigning 10 points to a piece of information means that it is the only one that is useful. Assigning 5 points to each piece of information means that they are equally useful. Keep in mind that **the total number of points assigned to A and B must add up to 10**.
>
> ________ A. Of the elderberries in that area, 60% are black elderberries and 40% are red elderberries.
>
> ________ B. When shown photographs of the two kinds of elderberries, Martha was able to correctly identify each of the varieties 20% of the time and incorrectly identified them 80% of the time.
>
> (Check your answer to make sure the numbers on Line A and Line B add up to 10.)

PART II

Now, on the basis of the data given above ("A" and "B"), estimate the likelihood that the berries Martha picked were black elderberries by entering a number **between 0 and 100**, where "0" means that they definitely were not black elderberries (that is, they were really red), and "100" means that they definitely were black elderberries.

What is the likelihood that the berries were black elderberries?______

Procedure. The participants were tested in their classrooms. Each participating student received a sheet of paper with one question on it; there was sufficient space to carry out calculations if a student wished to do so. The students read the scenario and answered Part I, in which they made judgments of the relative importance of two pieces of information; after completing Part I, they answered Part II, in which they made a single likelihood estimate. After completing the task, the participants returned their papers to the experimenters.

Results and Discussion

An ANOVA performed on the likelihood estimates showed significant effects of base rate and witness accuracy [for base rate, $F(3,350) = 39.79$; for witness accuracy, $F(2,350) = 40.17$, $MS_e = 505.71$; both ps $< .01$]. There was no significant effect of the interaction of base rate and witness accuracy [$F(6,350) = 1.76$]. Variance accounted for, ω^2, was 20.7% for base rate and 13.9% for witness accuracy. Median estimates, along with the estimates calculated using Bayes's theorem, appear in Figure 4. As has been seen in the other experiments, the participants tended to overestimate likelihood at lower levels of base rate and witness accuracy and underestimate it at higher levels. The mean amount of error in the students' estimates, defined as the absolute value of the deviation from the Bayesian answer, was 21.96 ($SE = 1.15$). An ANOVA revealed that there was a significant effect of base rate on amount of error [$F(3,350) = 13.37$]; there was also a significant effect of witness accuracy [$F(2,350) = 9.99$] and a significant base rate $\times$ witness accuracy interaction [$F(6,350) = 5.92$, $MS_e = 388.94$; all ps $< .01$]. Student–Newman–Keuls tests showed that there was significantly more error for estimates involving BR 5 than for those involving BRs 40, 60, and 95, which did not differ from one another; there was significantly less error for estimates involving WA 50 than for those involving WAs 20 and 80, which did not significantly differ. The average deviation from the Bayesian answer for each base-rate/witness-accuracy combination appears in Table 4.

When estimates were scored according to whether they were exact Bayesian responses, it was found that, overall, 16.8% of the answers met this criterion. There was no significant effect of base rate, according to a χ^2 analysis. However, there were significantly more Bayesian answers to WA 50 questions (34.2%) than to WA 20 (7.6%) or WA 80 (8.9%) questions [$\chi^2(2, N = 362) = 38.48$, $p < .01$].

What weights did the participants assign to base rate and witness accuracy information, and was it in any way related to their likelihood estimates? Since base rate and witness accuracy shared 10 points that the participants could allocate as they wished, any gain in the rating of one necessarily implies a loss in the rating of the other; therefore, it is necessary only to present the results for one factor. The overall mean weight assigned to base rate was 4.38 out of 10 possible points. However, as is shown in Figure 5, there was a great deal of individual variation in the weights assigned to base rate and witness accuracy. An ANOVA was performed to evaluate the effects of base rate and witness accuracy levels on the weights assigned to base rate as a source of information (WtB-R). The ef-

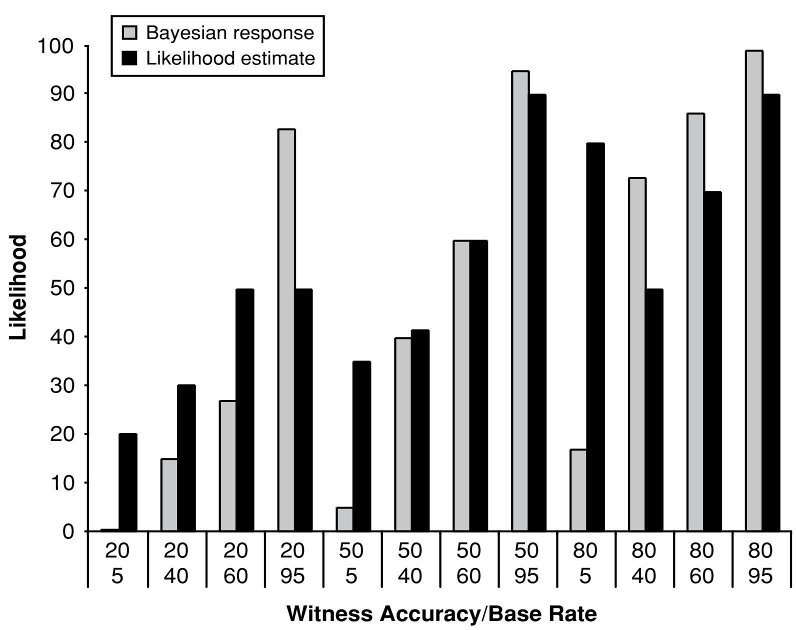

Figure 4. Experiment 5: Median likelihood estimates and corresponding Bayesian estimates.

Table 4
Differences Between Judged Likelihoods and Bayesian Likelihoods Experiment 5

	Witness Accuracy							
	20		50		80			
Base Rate	*M*	*SE*	*M*	*SE*	*M*	*SE*	Mean	*SE*
			Absolute Values of Differences					
5	25.58	3.30	29.74	4.98	46.43	5.15	34.17	3.09
40	21.13	1.82	11.39	1.95	22.32	2.65	18.25	1.61
60	24.43	3.29	6.86	1.76	19.15	3.20	17.11	1.82
95	35.08	4.96	15.44	4.31	11.73	1.65	20.31	2.49
Mean	25.79	2.07	15.37	1.80	24.70	2.00	21.96	1.15
			Directional Differences					
5	25.58	5.22	29.37	5.07	42.64	6.25	32.74	3.28
40	19.71	3.53	4.03	2.62	−19.73	3.17	−1.52	2.35
60	20.70	4.07	−3.14	2.11	−17.58	3.47	−0.52	2.56
95	−27.33	6.75	−14.63	4.42	−11.73	1.65	−17.61	2.74
Mean	11.69	2.96	3.86	2.26	−3.39	2.97	3.92	1.62

fect of base rate on WtB-R was not statistically significant [$F(3,355) = 2.48, p = .06$]. There was a significant effect of witness accuracy on WtB-R [$F(2,355) = 8.76, MS_e = 6.26, p < .01$]; a Student–Newman–Keuls test ($\alpha = .05$) showed that the importance of base rate information was rated significantly higher ($M = 5.15$, $SE = 0.25$) for the WA 50 question than for the WA 20 ($M = 4.0$, $SE = 0.23$) or the WA 80 ($M = 3.98$, $SE = 0.21$) question. It is notable that it was only when the witness information was completely unreliable that base rate information received more than half of the available weighting; most of the time, information about witness accuracy was rated as more important. There was no significant effect on WtB-R of the interaction of base rate and witness accuracy [$F(6,355) = 1.40$].

The students' weightings for base rate and witness accuracy information were also examined as a categorical variable, WtCat, with three levels: base rate and witness accuracy weighted equally, base rate weighted more than witness accuracy, and witness accuracy weighted more than base rate. Overall, 18.8% of the participants weighted base rate and witness accuracy equally, 30.2% gave more weight to base rate, and 51% gave more weight to witness accuracy. These response distributions were affected by base rate [$\chi^2(6, N = 367) = 20.93$] and by witness accuracy [$\chi^2(4, N = 367) = 20.74$; both ps $< .01$]. With respect to the effect of base rate, the participants who received questions with extreme base rates (BR 5 and BR 95) tended to assign more weight to base rate information than to witness information; the participants with BR 40 and

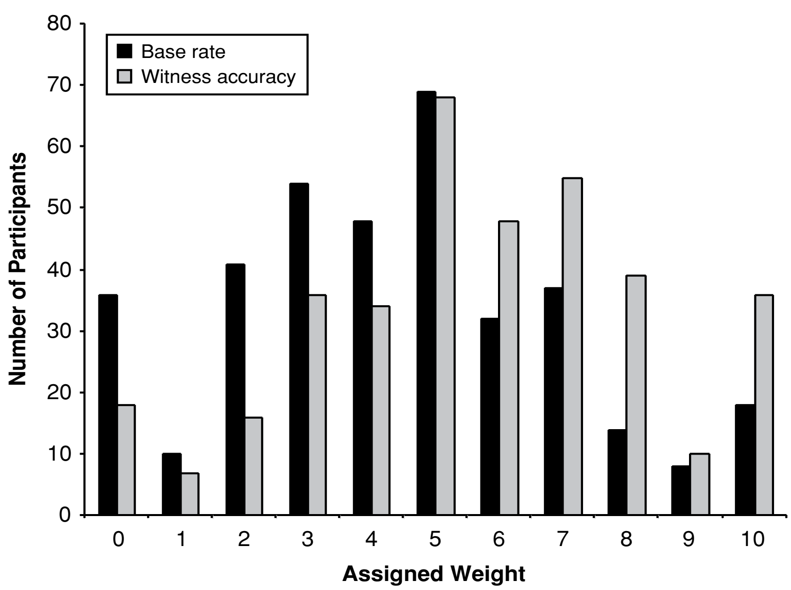

Figure 5. Experiment 5: Distribution of weights given to base rate and witness information.

BR 60 questions were more likely to assign more weight to witness information. With respect to witness accuracy, the participants answering WA 50 questions tended to assign more weight to base rate information than to witness information; those with WA 20 and WA 80 questions were more likely to assign the most weight to information about the witness.

Finally, were students' relative weighting of base rate and witness accuracy related to whether they gave exact Bayesian likelihood estimates? For the 362 participants for whom both pieces of information were available, this relationship was tested. Of the participants who gave more weight to base rate information, 27.8% gave Bayesian answers, in comparison with 13.4% for those who gave more weight to witness information and 11.9% for those who rated them equally. This effect was statistically significant [$\chi^2(2, N = 362) = 11.59, p < .01$]. However, because most of the participants who gave more weight to base rate information were also those who had received WA 50 questions, it is likely that this was the source of the effect.

GENERAL DISCUSSION

Did participants in the present study ignore base rate information in making their likelihood estimates? The evidence from the within-subjects experiments (Experiments 1 and 2) shows that the participants' estimates were responsive to changing base rates, as well as to changing witness accuracies. As has been noted by other investigators (see, e.g., Peterson & Beach, 1967), estimates tended to be lower than the normative response when likelihood was high and higher than the normative response when likelihood was low. However, as can be seen in Figure 1, when given multiple trials, the participants made likelihood estimates that increased regularly as base rates increased. Did the participants treat base rate information as less important than case cue information? The evidence from Experiments 3, 4, and 5, shown in Figures 2–4, is that when asked to make only a single judgment, they did. In this case, their estimates were less responsive to base rates and, as can be seen in Tables 2–4, were further, on average, from the estimates predicted by Bayes's theorem, in comparison with the within-subjects results of Experiment 1. Also, when the Experiment 5 participants were asked to give relative weighting to the importance of base rate and case cue (witness) information, they typically rated the case cue information as more important. But when the salience of base rate information was increased by varying base rates across trials, the participants clearly did not ignore base rates.

Another way in which base rate information could be made more salient was by describing the witness as unreliable—that is, as correct only 50% of the time. In this case, the participants' likelihood estimates were more accurate than they were for other questions. Not all the participants ignored the witness information and relied exclusively on the base rate when the witness was described as completely unreliable. Some may have thought that since both base rates and witness information had been provided by the experimenters, both should be utilized. Others may have failed to take into account that, for this task, 50% witness accuracy meant complete unreliability. Since in daily life many choice situations include more than two options, a 50% level of accuracy often does not imply that case cue information ought to be ignored. Finally, there is evidence that participants are influenced by any case cue information, even if it is of dubious value (Goodie & Fantino, 1995). Still, the results showed that at least some of the participants attended exclusively to base rates when they were the only reliable source of information available. Thus, making base rates more salient by neutralizing competing information made the participants' likelihood estimates more accurate.

Even when the participants recognized the need to consider both base rates and witness accuracies in estimating likelihood, they still faced the challenge of integrating these sources of information. The ease of integration was affected by the relative strength of the two sources of information. The participants' estimates were especially far from the Bayesian estimate when there was a large discrepancy between base rate and witness accuracy. This trend was evident in Experiment 1, as can be seen in Table 1; it was even more evident in the three between-subjects experiments. In part, this tendency toward greater error of estimation may occur because there is greater room for error on these judgments than on ones in which the base rate and case cue values are closer to one another. However, it likely was more difficult for the participants to integrate information from sources that appeared to contradict one another than when the data supplied were more congruent. For example, Figures 2–4 show that, when base rates were low, the participants' median estimates tended to match the witness reliability; this is consistent with findings that participants overestimate the likelihood of a person's having a rare disease, given a positive test result (see Gigerenzer, 1998, for examples).

Although many of the participants' estimates were far from the Bayesian estimate, some conformed to it closely or exactly. About half of the estimates made by the participants in Experiment 1 were within 10 points on either side of the Bayesian estimate; some of the participants made a large number of exact Bayesian responses. In the between-subjects experiments, the percentage of exact Bayesian estimates ranged from 12.25% in Experiment 3 to 16.8% in Experiment 5. There was no consistent trend across experiments showing how base rate magnitude affected Bayesian responding. However, the percentage of participants making Bayesian estimates varied according to the level of witness accuracy; in each case, there were more exact Bayesian estimates when witness accuracy was reported to be 50%, where the Bayesian answer was the same as the base rate. It is worth noting that for witness accuracies greater or less than 50%, few participants gave estimates corresponding to the base rate; thus, the larger number of Bayesian estimates for the WA 50 questions does not reflect a general tendency for participants to match their estimates to the base rate. The results of Experiment 3 demonstrated no effect of witness type, animate or inanimate, on the participants' responses.

The present work does not make comparisons between scenarios presented in likelihood and in frequency formats. However, the follow-up questions given to the participants in Experiment 2 showed rates of correct answers comparable to those reported by Gigerenzer and Hoffrage (1995), who reported 50% correct responses in the natural frequency format, in comparison with 28% correct responses in a probability format. In the Experiment 2 follow-up, we found (including both our trained participants and the untrained comparison group) 24% correct responses (within 5 percentage points of the Bayesian answer) to the likelihood format question and 48% correct responses (the exact Bayesian answer) to the natural frequency format question.

Experiments 1 and 2 showed small positive effects of feedback. However, even with the fairly intensive training given in Experiment 2, there was no evidence that the participants developed an accurate picture of the way in which base rate and case cue information interact to determine Bayesian likelihood estimates. In these experiments, we made no attempt to train the participants to calculate the Bayesian estimates; indeed, they had no real opportunity to do so. However, we anticipated that they might learn certain patterns—for example, that when base rate is 20 and witness accuracy 30, the likelihood of an outcome is less than 20, or that when base rate is 80 and witness accuracy is 70, the likelihood is greater than 80. Although the participants learned these patterns on a short-term basis, it does not appear that they understood their implications; thus, they did not generalize their learning effectively. This is consistent with the findings of Christensen-Szalanski and Beach (1982) and Lindeman, Van den Brink, and Hoogstraten (1988), who also found little generalization after their training procedures. Similarly, in a study of the conjunction effect, Stolarz-Fantino, Fantino, Zizzo, and Wen (2003, Experiment 5) found that for participants given feedback, monetary rewards, or both, the effect occurred at rates similar to those for control participants. Given the large number of trials required to induce the participants in Goodie and Fantino's studies to utilize base rates, perhaps we should not be surprised. Possibly, hundreds of trials with the present questions would result in better learning, as in Goodie and Fantino (1999b). It is also possible that calling participants' attention to the relevant variables during the task instructions would improve performance.

One way in which the repeated trials with feedback in Experiment 2 were effective was in increasing the participants' awareness of the (theoretical) importance of both base rate and case cue information in making estimates, as demonstrated by their written protocols. Lindeman et al. (1988) also found that base rates were mentioned more often in participants' protocols after training. They reported finding a great deal of individual variation in their participants' opinions as to whether base rate information was important; such variability was also evident in the weights given to base rate and witness information in the present Experiment 5, as is evident in Figure 5. As was noted by Donahoe and Palmer (1994), "in complex decision making, or judgment, behavior is guided by the combined effects of many stimuli, and there may be differences among decision makers in the stimuli to which they 'attend'" (p. 157).

What influences one participant to weight base rate and case cue information one way, whereas another evaluates it differently? In the absence of obvious instructions provided by the experimenters, participants can rely only on their own past histories to provide them with a context for the task (Fantino & Stolarz-Fantino, 2003). Thus, for example, they are predisposed to attend to information that varies, as it does in the within-subjects presentation (see, e.g., Kahneman, 2003). Their past experience also affects their interpretation of the task requirements, influencing whether they base their judgments on impressions or on deliberate reasoning, as described in dual-process models of cognition (e.g., Kahneman, 2003; Shafir & LeBoeuf, 2002; Stanovich, 1999; Stanovich & West, 2000). These models assume the operation of two systems: System 1, which is associative, automatic, and contextualized, and System 2, which is rule-based, analytic, and less dependent on context. Participants' construal of a task should influence which system is primarily activated. In Experiments 1 and 2, the participants viewed the scenarios on computer screens and made their estimates without benefit of calculators or, usually, even paper and pencils. Thus, it is likely that they construed the task as requiring them to form impressions, rather than perform calculations. What is not clear is how the participants in the between-subjects experiments construed the task. Some provided exact responses, especially when it was easy to do, as when the witness was unreliable or the base rate and witness accuracy were close in magnitude. But the diversity of responses to the task may reflect diversity in how it was viewed by the participants.

Goodie and Fantino (1996) found that participants' past history of reinforcement for matching led them to neglect base rate information and attend to case cue information on a matching-to-sample analogue of the taxicab problem. In this behavioral task, base rate neglect resulted from preexisting associations between stimuli; the participants had learned to pay attention to case cues. Pigeons in the same task selected optimally, suggesting that the relatively "uneducated" pigeon is aptly sensitive to the relevant case cues and base rates (Hartl & Fantino, 1996). However, when pigeons were given extensive training with the matching-to-sample task, they too neglected base rate information (Fantino, Kanevsky, & Charlton, 2005), underscoring the role of experience in base rate neglect. These considerations also apply to the statistical form in which most base rate problems are presented in the laboratory (Stolarz-Fantino & Fantino, 2005). In daily life, both base rates and case cue accuracies are learned through experience, one instance at a time. However, as has been pointed out by Fiedler (2000), the base rates of many important events are, for the most part, unknown. So, for example, it may be easy to learn, by examining accident statistics, how likely it is that people who have automobile accidents have been drinking; however, it is harder to know the base rate of drinking among drivers and, thus, the likelihood that a drinking driver will have an accident. Just as Goodie and Fantino's (1996) participants had extensive experience with matching, people usually have more experience with case cue information than with base rates. Base rates are a less direct source of information, in the sense that, in order to be influenced by them, the decision maker needs to integrate events occurring over a period of time; a case cue, on the other hand, is fully present at the time of the decision. In a paper-and-pencil task, participants receive base rate and case cue information simultaneously, in statistical form; however, they may be predisposed by past experience to favor the case cue.

In summary, the participants in the present study made more effective use of base rate and case cue information when the design of the experiment caused these factors to vary, as they do in multiple-trial experiments. Estimates were also more accurate when base rates and case cues were congruent and in situations in which it was normative to judge on the basis of only one piece of information because the other was unreliable. Even under testing conditions unfavorable to performing calculations, many participants' estimates were close to the Bayesian estimate, and some conformed to it exactly. Thus, although base rate neglect was evident throughout these experiments, base rates were not ignored. Experience predisposes participants to neglect base rates. However, when base rates are made salient by experimental manipulation, base rate neglect is minimized or eliminated.

REFERENCES

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

Birnbaum, M. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, **96**, 85-93.

Birnbaum, M., & Mellers, B. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality & Social Psychology*, **45**, 792-804.

Case, D. A., Fantino, E., & Goodie, A. S. (1999). Base-rate training without case cues reduces base-rate neglect. *Psychonomic Bulletin & Review*, **6**, 319-327.

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

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

Cohen, L. J. (1979). On the psychology of prediction: Whose is the fallacy? *Cognition*, **7**, 385-407.

Cohen, M. S. (1993). Three paradigms for viewing decision biases. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 36-50). Norwood, NJ: Ablex.

Doherty, M. E. (2003). Optimists, pessimists, and realists. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 643-679). Cambridge: Cambridge University Press.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Boston: Allyn & Bacon.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). Cambridge: Cambridge University Press.

Fantino, E., Kanevsky, I. G., & Charlton, S. (2005). Teaching pigeons to commit base-rate neglect. *Psychological Science*, **16**, 820-825.

Fantino, E., & Stolarz-Fantino, S. (2003). Context and its effect on transfer. *Greek Economic Review*, **22**, 11-26.

Fiedler, K. (2000). Beware of samples! A cognitive–ecological sampling approach to judgment biases. *Psychological Review*, **107**, 659-676.

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

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, **103**, 684-704.

Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counseling for low-risk clients. *AIDS Care*, **10**, 197-211.

Goodie, A. S., & Fantino, E. (1995). An experientially derived base-rate error in humans. *Psychological Science*, **6**, 101-106.

Goodie, A. S., & Fantino, E. (1996). Learning to commit or avoid the base-rate error. *Nature*, **380**, 247-249.

Goodie, A. S., & Fantino, E. (1999a). Base rates versus sample accuracy: Competition for control in human matching to sample. *Journal of the Experimental Analysis of Behavior*, **71**, 155-169.

Goodie, A. S., & Fantino, E. (1999b). What does and does not alleviate base-rate neglect under direct experience. *Journal of Behavioral Decision Making*, **12**, 307-335.

Hartl, J. A., & Fantino, E. (1996). Choice as a function of reinforcement ratios in delayed matching to sample. *Journal of the Experimental Analysis of Behavior*, **66**, 11-27.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, **12**, 275-305.

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

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, **106**, 62-88.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, **58**, 697-720.

Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society A*, **154**, 75-81.

Kaye, D. H., & Koehler, J. J. (2003). The misquantification of probative value. *Law & Human Behavior*, **27**, 645-659.

Kennedy, M. L., Willis, W. G., & Faust, D. (1997). The base-rate fallacy in school psychology. *Journal of Psychoeducational Assessment*, **15**, 292-307.

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

Leon, M., & Anderson, N. H. (1974). A ratio rule from integration theory applied to inference judgments. *Journal of Experimental Psychology*, **102**, 27-36.

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

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

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: Beyond heuristics and frequency format explanations. *Organizational Behavior & Human Decision Processes*, **82**, 217-236.

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

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, **53**, 491-517.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences*, **23**, 645-726.

Stolarz-Fantino, S., & Fantino, E. (2005). The rules we choose by. *Behavioral Processes*, **69**, 151-153.

Stolarz-Fantino, S., Fantino, E., Zizzo, D., & Wen, J. (2003). The conjunction effect: New evidence for robustness. *American Journal of Psychology*, **116**, 15-34.

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

Ward, L. M. (1975). Heuristic use or information integration in the estimation of subjective likelihood? *Bulletin of the Psychonomic Society*, **6**, 43-46.

Weber, E. U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1151-1164.

(Manuscript received August 18, 2004;
revision accepted for publication April 20, 2005.)