Psychological Review
1999, Vol. 106, No. 2, 417–424

0033-295X/99/$3.00

How to Improve Bayesian Reasoning: Comment on Gigerenzer and Hoffrage (1995)

Barbara A. Mellers and A. Peter McGraw
Ohio State University

G. Gigerenzer and U. Hoffrage (1995) claimed that Bayesian inference problems, which have been notoriously difficult for laypeople to solve using base rates, hit rates, and false-alarm rates, become computationally simpler when information is presented with frequencies based on natural sampling. They made an evolutionary argument for the improved performance. The authors of the present article show that performance can improve with either probabilities or frequencies, depending on the rareness of the events and the type of information presented. When events are rare, probabilities are more difficult to understand than frequencies (i.e., 5 out of 1,000 vs. .005.). Furthermore, when the information is presented as joint and marginal events, nested sets become more apparent. Frequencies based on natural sampling have these desirable properties. The authors agree with Gigerenzer and Hoffrage that frequencies can improve Bayesian reasoning, but they attribute that improvement to the use of mental models that involve elements of nested sets.

For years, psychologists have been interested in the psychological processes that describe how people assess uncertainty and change their opinions in light of new information. Edwards (1968; Phillips & Edwards, 1966) was one of the first to examine whether psychological processes could be represented with Bayes Theorem, a normative rule for revising beliefs in the face of new evidence. He concluded that people were "conservative" Bayesians who changed their views in the right direction but not to the right degree. Later, Kahneman and Tversky (1972) began a series of studies on probability inference and concluded that "man is apparently not a conservative Bayesian; he is not Bayesian at all" (p. 450). Some of the evidence used to support their claim came from the following problem about cabs (Tversky & Kahneman, 1982):

> 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? (p. 156–157)

Bayes Theorem is the normative rule for belief revision in the cab problem.[1] Let H and $-H$ represent the hypotheses that the cab involved in the accident is Blue or Green. Let D represent the data that bear on the hypotheses, namely, the witness's report of a Blue cab. The Bayesian solution, or the posterior probability of H given D, can be written:

$$p(H|D) = [p(D|H)p(H)]/[p(D|H)p(H) + p(D|-H)p(-H)], \quad (1)$$

where $p(H|D)$ is the posterior probability that the cab involved in the accident was Blue given the witness's report that it was Blue, $p(H)$ and $p(-H)$ are the base rates for Blue and Green cabs, respectively, and the conditional probabilities, $p(D|H)$ and $p(D|-H)$, are the witness's hit and false-alarm rates, respectively. Inserting the values into Equation 1 gives the following:

$$[(0.8)(0.15)]/[(0.8)(0.15) + (0.2)(0.85)] = 0.41.$$

It is not surprising that most people find this problem difficult. The median and modal response is usually 0.8, the witness's hit rate. Birnbaum and Mellers (1983) found that, after the hit rate, the next most frequent response is 0.15, the base rate, followed by 0.12, the hit rate times the base rate. People do not seem to know whether they should combine probabilities, and if so, how. Consequently, they tend to respond with a single probability or multiply them.

Tversky and Kahneman (1982) argued that results were consistent with the representativeness heuristic; most people make prob-

Barbara A. Mellers and A. Peter McGraw, Department of Psychology, Ohio State University.

Correspondence concerning this article should be addressed to Barbara A. Mellers, Department of Psychology, Ohio State University, Columbus, Ohio 43210. Electronic mail may be sent to mellers.1@osu.edu or mcgraw.27@osu.edu.

[1] Birnbaum (1983) argued that the normative analysis of the cab problem is incomplete. Problems of this type need a theory of the witness's behavior before a correct solution can be found. He showed how signal-detection theory provides a framework for entertaining a number of psychological theories of the witness.

ability judgments based on the extent to which an event A is representative of, or similar to, a class B. The base rate is less vivid and less compelling than the report, so people focus on the witness's testimony. This interpretation has generated a great deal of research and many lively debates (see Gigerenzer & Murray, 1987; Koehler, 1996).

Gigerenzer and Hoffrage (1995) offerred a new perspective (also developed by Cosmides and Tooby, 1996) that people, although not Bayesians per se, can reason more accurately when information is presented as frequencies based on natural sampling. Natural sampling is defined as "the sequential acquisition of information by updating event frequencies *without* artificially fixing the marginal frequencies" (Gigerenzer & Hoffrage, 1995 p. 686). They illustrated the idea with a story:

> Imagine an old, experienced physician in an illiterate society. She has no books or statistical surveys and therefore must rely solely on her experience. Her people have been afflicted by a previously unknown and severe disease. Fortunately, the physician has discovered a symptom that signals the disease, although not with certainty. In her lifetime, she has seen 1,000 people, 10 of whom had the disease. Of those 10, 8 showed the symptom; of the 990 not afflicted, 95 did. Now a new patient appears. He has the symptom. What is the probability that he actually has the disease? (p. 686)

In this problem, the physician has evaluated the diagnosticity of the symptom in all members of the population, so the marginal frequencies of disease and no disease used in the test are identical to those naturally occurring in the population. Natural sampling differs from systematic sampling in which the marginal frequencies differ from the frequencies in the population. For example, if the symptom was tested on a subset of the population (e.g., 10 people with the disease and 10 people without), the sampling would be systematic, not natural.

Gigerenzer and Hoffrage (1995) argued that the mind can handle frequencies based on natural sampling because this sequential way of acquiring information has occurred throughout human evolution. People come across events, one by one, and add them to the appropriate mental set. The physician in the medical problem keeps track of instances in which both the symptom and the disease occur relative to instances in which the symptom occurs, regardless of the disease.

To test their claim that frequencies based on natural sampling improve Bayesian reasoning, Gigerenzer and Hoffrage (1995) conducted two tests. First, they presented people with information using Equation 1, which they called the standard menu. Some participants were given either probabilities or frequencies based on natural sampling. For example, in the mammography problem (taken from Eddy, 1982), Gigerenzer and Hoffrage told participants who received probabilities that the probability of breast cancer for women at age 40 is 1% (base rate); of those who have breast cancer, the probability of a positive mammogram is 80% (hit rate); and of those without breast cancer, the probability of a positive mammogram is 9.6% (false-alarm rate). Participants given frequencies based on natural sampling were told that 10 of 1,000 women have breast cancer at age 40. Of those 10, 8 have positive mammographies (hit rate), and of the 990 without breast cancer, 95 have positive mammographies (false-alarm rate). Gigerenzer and Hoffrage found that performance in this problem and similar problems was superior with frequencies based on natural sampling. Overall, 46% of responses were correct with frequencies based on natural sampling, and only 16% were correct with probabilities.

In the second test, Gigerenzer and Hoffrage (1995) presented participants with information using another form of Bayes Theorem. The joint probability of the hypothesis and the data can be written:

$$p(H \,\&\, D) = p(D|H)p(H),$$

so Bayes Theorem can be expressed as:

$$p(H|D) = [p(H \,\&\, D)]/[p(H \,\&\, D) + p(-H \,\&\, D)], \quad (2)$$

where $p(H \,\&\, D)$ is the joint probability of the hypothesis and the data, and $p(-H \,\&\, D)$ is the joint probability of the counter hypothesis and the data. Because there are only two ways that

Table 1
Problems Used by Gigerenzer and Hoffrage (1995) in Experiment 1

	Standard information			Short information	
Problem	$p(H)$	$p(D\|H)$	$p(D\|-H)$	$p(H \& D)$	$p(D)$
1. Breast cancer	.01000	.800	.0960	.008000	.10300
2. Prenatal damage	.00210	.480	.0050	.001000	.00600
3. Blue cab	.15000	.800	.2000	.120000	.29000
4. AIDS	.00010	1.000	.0010	.000100	.00100
5. Heroin addict	.00010	1.000	.0020	.000100	.00200
6. Pregnant	.02000	.950	.0050	.019000	.02400
7. Car accident	.01000	.550	.0500	.005500	.05600
8. Bad posture	.05000	.400	.2000	.020000	.21000
9. Accident	.03000	.900	.4000	.027000	.42000
10. Suicide	.00024	.150	.1200	.000036	.12000
11. Red ball	.80000	.750	.2500	.600000	.65000
12. Economics course	.30000	.700	.5000	.210000	.56000
13. Active feminist	.05000	.004	.0210	.000200	.02000
14. Pimp	.00005	.800	.0005	.000040	.00054
15. Admissions	.36000	.750	.2000	.270000	.40000

Table 2
Percentage of Participants Using Different Strategies in Study 1

		Probabilities		Frequencies			
				Sys sampling		Nat sampling	
Strategy	Formal equivalence	Stand	Joint	Stand	Joint	Stand	Joint
Bayesian solution	$p(B\|P) = .08$	7	3	20	16	28	21
Base rate	$p(B) = .01$	17	10	23	16	17	19
Hit rate	$p(P\|B) = .8$	5	0	9	0	7	0
False-alarm rate	$p(P\|-B) = .096$	2	5	0	0	0	7
Joint probability	$p(P \& B) = .008$	12	46	17	23	13	21
Probability of data	$p(P) = .103$	5	3	3	16	7	2
Sample size (*n*)		42	39	35	44	46	42

Note. Sys = systematic; Nat = natural; Stand = standard.

the data can occur, either with or without H, the sum of $p(H \& D)$ and $p(-H \& D)$ is $p(D)$, so Bayes Theorem can be expressed as:

$$p(H|D) = [p(H \& D)]/[p(D)]. \quad (3)$$

Equation 3 shows the form of Bayes Theorem that Gigerenzer and Hoffrage (1995) used in their second test. They called this type of information the *short menu.* Participants received either probabilities or frequencies based on natural sampling. For the mammography problem, participants given probabilities were told that the probability of breast cancer and a positive mammography for women at age 40 is 0.8%, and the probability of a positive mammography for women at age 40 is 10.3%. Participants given frequencies based on natural sampling were told that, of 1,000 women at age 40, 8 have breast cancer and positive mammographies and 103 have positive mammographies. Once again, Gigerenzer and Hoffrage found that performance was better with frequencies based on natural sampling. With frequencies, 50% of the responses were correct, and with probabilities, only 28% were correct.

Reasoning With Elements of Nested Sets

We believe that frequencies based on natural sampling improve performance in Bayesian inference problems because, with frequencies, rare events are easier to comprehend than with probabilities. For example, most untutored people lack an intuitive feel for the difference between 0.0005 and 0.00005, but this difference is easier to understand when expressed with frequencies as 5 in 10,000 and 5 in 100,000. Furthermore, frequencies based on natural sampling are advantageous because they help people visualize nested sets, or subsets relative to larger sets.

In their tests, Gigerenzer and Hoffrage (1995) presented participants with several problems, including the mammography problem (Eddy, 1982), the cab problem (Tversky & Kahneman, 1982), and a short version of Ajzen's (1977) economics problem. Table 1 lists problems and probabilities in the standard and short menus used by Gigerenzer and Hoffrage. In all but 4 of the 15 problems (3, 11, 12, and 15), one or more of the probabilities is relatively small (.05 or lower), and probabilities get as small as .000036. Frequencies are particularly helpful with these types of problems.

Table 3
Percentage of Participants Using Different Strategies in Study 2

		Probabilities		Frequencies			
				Sys sampling		Nat sampling	
Strategy	Formal equivalence	Stand	Joint	Stand	Joint	Stand	Joint
Bayesian solution	$p(H\|D) = .41$	4	16	5	—	8	17
Base rate	$p(H) = .15$	11	3	16	—	21	6
Hit rate	$p(D\|H) = .8$	32	2	30	—	14	1
False-alarm rate	$p(D\|-H) = .2$	1	1	5	—	5	1
Hit − false-alarm rate	$p(D\|H)-p(D\|-H) = .6$	5	3	8	—	1	0
Joint probability	$p(H \& D) = .12$	4	30	4	—	8	25
Joint probability	$p(-H \& D) = .17$	0	0	1	—	2	1
Probability of data	$p(D) = .29$	0	3	0	—	8	2
Maximum uncertainty	.5	8	14	5	—	3	11
Strategies not identified		35	26	24	—	30	43
Sample size (*n*)		122	96	138	—	131	139

Note. Sys = systematic; Nat = natural; Stand = standard.

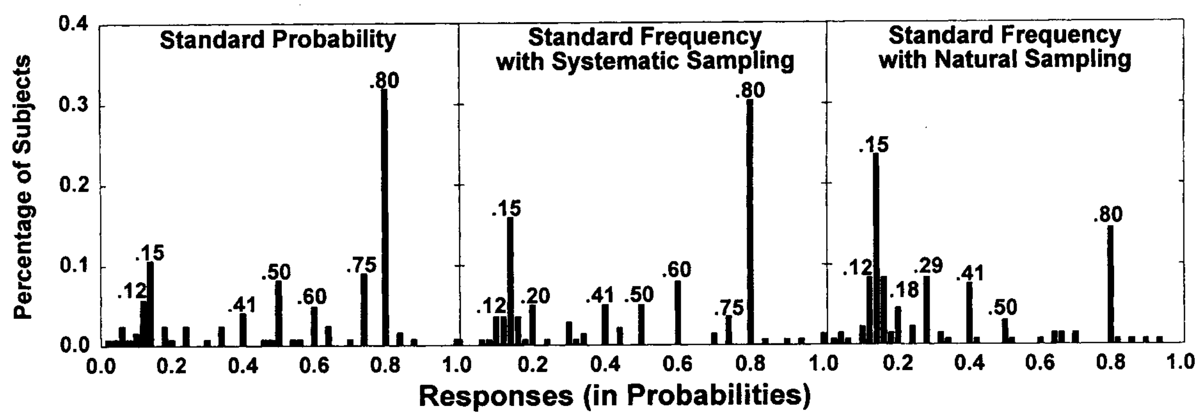

Figure 1. Percentage of participants giving responses plotted against responses (in probabilities). The left panel shows the distribution of responses from the standard probability task. The center and right panels show data from the standard frequency task with systematic and natural sampling, respectively. Responses are pooled over the three problems.

Gigerenzer and Hoffrage (1995) presented participants with information using the standard and short menus (Equations 1 and 3, respectively). The standard menu with probabilities does not lend itself easily to mental models with nested sets; most people cannot envision nested sets with $p(H)$, $p(D|H)$, and $p(D|-H)$. However, with natural sampling, joint events are easy to extract. People can compare the rate of H & D relative to D, which is the sum of H & D plus $-H$ & D. Equation 3, called the *short menu,* is even easier because $p(H$ & $D)$ and $p(D)$ are provided. This type of information should improve performance with both frequencies based on natural sampling and probabilities. However, improvement with probabilities would be smaller if events were rare and the differences between probabilities were harder to understand.

We now examine whether these properties of frequencies based on natural sampling improve performance. We predict that, with rare events, performance should be better with frequencies than probabilities. Furthermore, with the standard menu, performance should be better with frequencies based on natural sampling than with probabilities because people can envision nested sets. Performance with frequencies based on natural sampling should be better than performance with systematic sampling because, with systematic sampling, frequencies of H, $D|H$, and $D|-H$ do not easily bring to mind mental models of nested sets.

We also compare frequencies with probabilities using with another form of Bayes theorem called the joint menu (Equation 2). With frequencies based on natural sampling, this type of information also allows people to visualize nested sets. When events are rare, performance with probabilities should be worse than that with frequencies based on natural sampling.

When events are more common, the advantage provided by frequencies should be minimized. With the standard menu, performance using frequencies based on natural sampling should be better than that with probabilities. With the joint menu, there should be no difference in performance. We now examine these predictions.

Study 1: Rare Events

We provided 248 people with different versions of the mammography problem (Appendix A). Each person received one problem. Six versions of the problem were constructed from a 2×3 design of Information Type (standard or the joint menu) $\times$ Information Form (probabilities, systematic sampling of frequencies, or natural sampling of frequencies). Undergraduates at the University of California, Berkeley, who were taking upper and lower division psychology courses volunteered to participate when the experimenter visited their classroom and asked for their assistance. Participants were told that they could use pencils, paper, and calculators. They were also asked to explain their reasoning. The task took approximately 10 min.

Table 2 shows percentages of different responses. The percentage of Bayesian solutions for different tasks appears in the first row. Performance was poor when participants were given probabilities for both standard and joint menus; only 7% and 3% of participants in the standard and joint tasks gave Bayesian solutions, respectively. In the standard task, there were numerous strategies, and no single response dominated. In the joint task, almost half of the participants gave the joint probability, $p(B$ & $P)$, as their response. Either they confused $p(B$ & $P)$ with $p(B|P)$, or they made arithmetic errors (.008 vs. .08) when trying to reason with small probabilities.

When information was given in the form of frequencies based on natural sampling, participants did better in both the standard and the joint tasks. With natural sampling, participants were correct 28% and 21% of the time, respectively. With systematic sampling, percentages were 20% and 16% in the standard and joint tasks, respectively. Systematic sampling resulted in performance that was slightly worse than natural sampling, presumably because nested sets are not as easy to visualize with systematic sampling.

In summary, Study 1 shows that, when problems contain rare events, the extent to which people reason in a Bayesian fashion depends to a greater extent on the form of the information than the type of the information. Participants gave more accurate responses with frequencies than probabilities, for both standard and joint tasks. We now examine inference problems that do not contain rare events.

Study 2: More Common Events

In this study, we gave 626 people 1 of 15 problems. We used five versions of the cab problem (Tversky & Kahneman, 1982), the

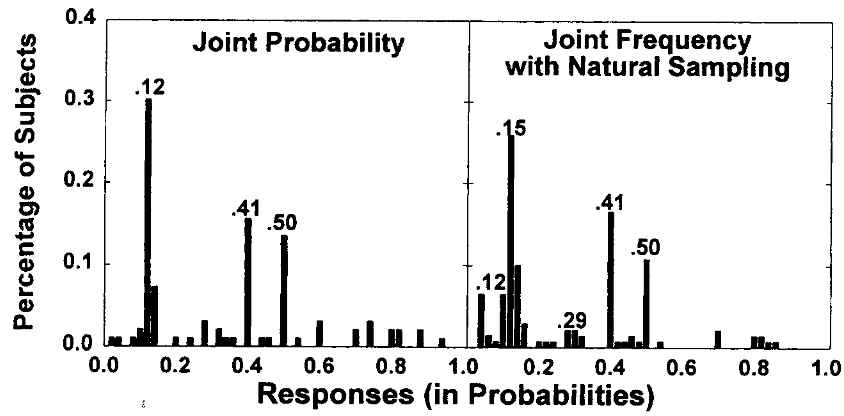

Figure 2. Percentage of people giving different responses in the joint probability task and the joint frequency task with natural sampling, plotted as in Figure 1.

used car problem (Birnbaum & Mellers, 1983), and the light bulb problem (Lyon & Slovic, 1976). To minimize extraneous variance, we modified the used car problem and the light bulb problem so that probabilities were identical to those in the cab problem. The five versions of the cab problem are shown in Appendix B. These versions were similar to those in Study 1, except we did not include a joint frequency task based on systematic sampling.[2] Results were similar for each of the three problems, so we present only the pooled data below. Participants were students in an introductory psychology course at Ohio State University who received course credit for their participation.

The first row of Table 3 shows the percentages of correct responses pooled over the three inference problems with identical probabilities. With probabilities, performance depended greatly on the type of information provided. In the standard task, the percentage correct was 4%, but in the joint task, the percentage correct increased to 16%. With frequencies based on natural sampling, the percentage correct was 8% in the standard task and 17% in the joint task. One might wonder why performance in the standard frequency task was so poor. With the standard menu, people may not have recognized the nested sets. In sum, when inference problems contain larger probabilities, performance depends on the type of information to a greater extent than the form of the information.

Figure 1 shows the distribution of responses for the standard probability and the standard frequency tasks with systematic and natural sampling. Percentages of participants giving each response are plotted against responses (in probabilities). Distributions are remarkably similar for probabilities and frequencies based on systematic sampling. In both panels, the modal response was 0.8, commonly found in this problem. Participants tended to confuse the posterior probability, $p(H|D)$, with the hit rate, $p(D|H)$. Gigerenzer and Hoffrage (1995) also found that the modal response was 0.8 in their standard probability task with the cab problem.[3]

The distribution for frequencies based on natural sampling is shown in the third panel. The modal response was the base rate, not the hit rate, and the next most frequent response was the hit rate, followed by the joint probability, $p(H \& D)$. Gigerenzer and Hoffrage (1995) found that, with this problem, the most frequent response was $p(H \& D)$.

Figure 2 shows the distribution of responses for the joint probability task on the left and the joint frequency task with natural sampling on the right. Once again, distributions are very similar. The similarity of these distributions runs counter to the evolutionary explanation, but is consistent with the idea that, if events are more common, joint menus are advantageous because they help people to visualize nested sets.

In summary, when inference problems contain larger probabilities, performance depends on the type of information more than the form of the information. Neither probabilities nor frequencies consistently produce better performance. Frequencies in the joint task based on natural sampling resulted in a greater percentage of accurate responses than probabilities in the standard task. However, probabilities in the joint task produced a greater percentage of accurate responses than frequencies in the standard task.

Conclusion

In Bayesian inference tasks, people must assess uncertainty and revise their opinions in the face of new information. Such tasks are common in medical diagnoses (e.g., the appearance of a symptom that is correlated with a disease), legal settings (e.g., new evidence that bears on the innocence or guilt of the defendant), and scientific debates (e.g., data that are consistent with one theory and not another). People are not Bayesians, and most of the time, they find probabilistic inference tasks extremely difficult. Gigerenzer and Hoffrage (1995) showed that people can reason in a more Bayesian fashion when inference problems are presented with frequencies based on natural sampling. Frequencies have also been shown to improve reasoning with conjunctive problems, such as the heart attack problem (Tversky & Kahneman, 1982) and the Linda prob-

[2] With the joint task, natural sampling and systematic sampling are identical. They could differ if systematic sampling were based on fewer observations. For example, the problem might state that "on average, the witness correctly identified Blue cabs as Blue on 1.2 out of 10 tests. On average, the witness incorrectly identified Green cabs as Blue on 1.7 out of 10 tests."

[3] We thank Gigerenzer & Hoffrage for kindly providing their data for the cab problem from Study 1.

Table 4
Percentage of Participants With Bayesian Solutions in the Cab Problem

	Type of information			
	Gigerenzer and Hoffrage (1995)		Mellers & McGraw	
Form of information	Standard	Short	Standard	Joint
Probabilities	14	27	4	16
Frequencies: natural sampling	23	34	8	17

lem (Gigerenzer, 1991). In the Linda problem, participants are given a description of Linda, a philosophy major who is concerned with issues of social justice, and then they are asked which occupations and hobbies best describe Linda. When the problem is presented in a probability format as a description of one person, people tend to say that "bank teller and feminist" is more likely to describe Linda than "bank teller." This pattern is a clear-cut violation of the conjunction rule, which says that the probability of being a "bank teller and a feminist" cannot be greater than the probability of being a "bank teller." When the problem is presented in a frequency format and phrased in terms of 100 people who fit the description of Linda, "bank tellers" are judged as more likely than "bank tellers and feminists" (Gigerenzer, 1991).

Frequencies based on natural sampling improve Bayesian inference for two reasons. First, frequencies make rare events easier to understand, because information is represented as elements of a set. Second, in the standard menu, frequencies based on natural sampling allow people to visualize nested sets, extract joint events $H \& D$ and $-H \& D$, and compute the frequency of $H \& D$ relative to D. Other types of information also allow people to visualize nested sets, including information in Equation 2 or 3.

We demonstrate that when participants are given versions of the mammography problem, which contains rare events, performance depends largely on the form of the information. Performance is worse with probabilities, regardless of the menu. Accuracy improves when participants can visualize nested sets (i.e., in the standard and joint frequency tasks with natural sampling). When participants are given inference problems with more common events, performance depends largely on the type of information. The benefit provided from frequencies is not as great.

We compared our results in Study 2 with those obtained by Gigerenzer and Hoffrage (1995) in their Study 1 for the cab problem. Percentages of participants giving Bayesian answers in both studies are shown in Table 4. First, we compare responses in the standard task. Only 4% of participants gave the correct response in our study, whereas 14% gave the correct response in Gigerenzer and Hoffrage's study.[4] Next we compare responses in the joint task from our study with the short task from Gigerenzer and Hoffrage's study. Performance was better in the joint task or the short task than in the standard tasks. Perhaps more interesting is the comparison of the joint or short probability task with the standard frequency task. Both studies show a greater percentage of Bayesian responses with probabilities than frequencies. In sum, responses based on probabilities can be superior to those based on frequencies with natural sampling, depending on the type of information and rareness of the events.

In conclusion, if one had to select a single form in which to present information, we would recommend frequencies. With rare events, frequencies facilitate Bayesian reasoning. If one had to select a single type of information, we would recommend the short task or, more generally, any type of information that allows people to visualize nested sets. These types and forms of information help people make better inferences, although neither one is a silver bullet. Bayesian reasoning is hard for both experts and nonexperts, and there is plenty of room for additional improvement.

[4] Although these differences are troublesome, relative differences are similar in both studies. There may be several reasons that Gigerenzer and Hoffrage (1995) found more participants giving correct answers. Gigerenzer & Hoffrage gave their participants many problems, whereas we gave ours only one. Those who received several problems may have benefitted from practice effects. Furthermore, students at the University of Salzburg, Austria, may be better trained than the undergraduates we sampled at Ohio State University.

References

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

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

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

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

Eddy, D. M. (1982). Probability 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, England: Cambridge University Press.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmutz (Ed.), *Formal representations of human judgment* (pp. 17–52). New York: Wiley.

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

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

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

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

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

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

Phillips, L., & Edwards, W. (1966). Conservatism in a simple probability model inference task. *Journal of Experimental Psychology, 72,* 346–354.

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, England: Cambridge University Press.

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

Appendix A

Versions of the Mammography Problem

Standard Probability Task:

The probability of breast cancer is 1% for a woman at age 40 who participates in routine screening. The probability is 80% that a woman with breast cancer will get a positive mammography. The probability is 9.6% that a woman *without* breast cancer will get a positive mammography. A woman at age 40 has a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

_____%

Standard Frequency Task with Systematic Sampling:

The frequency of breast cancer is 1 out of every 100 women at age 40 who participate in routine screening. On average, 80 out of every 100 women with breast cancer will get positive mammographies. On average, 9.6 out of every 100 women *without* breast cancer will get positive mammographies. Here is a new representative sample of women at age 40 who got positive mammographies in routine screening. How many of them actually have breast cancer?

_____out of_____

Standard Frequency Task with Natural Sampling:

The frequency of breast cancer is 10 out of every 1,000 women at age 40 who participate in routine screening. Eight out of every 10 women with breast cancer will get positive mammographies. 95 out of every 990 women *without* breast cancer will get positive mammographies. Here is a new representative sample of women at age 40 who got positive mammographies in routine screening. How many of them actually have breast cancer?

_____out of_____

Joint Probability Task:

The probability of breast cancer is 1% for a woman at age 40 who participates in routine screening. The probability is 0.8% that a woman with breast cancer will get a positive mammography. The probability is 9.5% that a woman *without* breast cancer will get a positive mammography. A woman at age 40 has a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

_____%

Joint Frequency Task with Systematic Sampling:

The frequency of breast cancer is 1 out of every 100 women at age 40 who participate in routine screening. On average, 0.8 out of every 100 women will get a positive mammography and have breast cancer. On average, 9.5 out of every 100 women will get a positive mammography and *not* have breast cancer. Here is a new representative sample of women at age 40 who got positive mammographies in routine screening. How many of them actually have breast cancer?

_____out of_____

Joint Frequency Task with Natural Sampling:

The frequency of breast cancer is 10 out of every 1,000 women at age 40 who participate in routine screening. 8 out of every 1,000 women will get a positive mammography and have breast cancer. 95 out of every 1,000 women will get a positive mammography and *not* have breast cancer. Here is a new representative sample of women at age 40 who got positive mammographies in routine screening. How many of them actually have breast cancer?

_____out of_____

Appendix B

Versions of the Cab Problem

Standard Probability Task:

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. 15% of all cabs in the city are Blue and 85% are Green. On the night of the accident, a witness identified the cab as "Blue." The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs. When the cabs were really Blue, the witness said they were Blue 80% of the time. When the cabs were really Green, the witness said they were Blue 20% of the time. What is the probability that the cab involved in the hit-and-run accident was Blue?

_____%

Standard Frequency Task with Systematic Sampling:

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. Of every 100 cabs in the city, 15 are Blue and 85 are Green. On the night of the accident, a witness identified the cab as Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs. When the cabs were really blue, the witness said they were Blue in 80 out of 100 tests. When the cabs were really Green, the witness said they were Blue in 20 out of 100 tests. What are the chances that the cab involved in the hit-and-run accident was Blue?

_____of_____

Standard Frequency Task with Natural Sampling:

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. Of every 100 cabs in the city, 15 are Blue and 85 are Green. On the night of the accident, a witness identified the cab as Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Green cabs. When the cabs were really Blue, the witness said they were Blue in 12 out of 15 tests. When the cabs were really Green, the witness said they were Blue in 17 out of 85 tests. What are the chances that the cab involved in the hit-and-run accident was Blue?

_____of_____

Joint Probability Task:

A cab was involved in a hit-and-run accident at night. Two cab companies, the Blue and the Green, operate in the city. On the night of the accident, a witness identified the cab as Blue. The court tested the reliability of the witness under similar visibility conditions with Blue and Green cabs. In 12% of the tests, the witness said the cabs were Blue and the cabs really were Blue. In 17% of the tests, the witness said the cabs were Blue and the cabs really were Green. What is the probability that the cab involved in the hit-and-run accident was Blue?

_____%

Joint Frequency Task with Systematic Sampling:

A cab was involved in a hit-and-run accident at night. Two cab companies, the Blue and the Green, operate in the city. On the night of the accident, a witness identified the cab as Blue. The court tested the reliability of the witness under similar visibility conditions with Blue and Green cabs. In 12 out of 100 tests, the witness said the cabs were Blue and the cabs really were Blue. In 17 out of 100 tests, the witness said the cabs were Blue and the cabs really were Green. What are the chances that the cab involved in the hit-and-run accident was Blue?

_____of_____

Received May 21, 1998
Revision received August 11, 1998
Accepted August 12, 1998 ■

New Editors Appointed, 2000-2005

The Publications and Communications Board of the American Psychological Association announces the appointment of three new editors for 6-year terms beginning in 2000.

As of January 1, 1999, manuscripts should be directed as follows:

- **For Experimental and Clinical Psychopharmacology, submit manuscripts to Warren K. Bickel, PhD, Department of Psychiatry, University of Vermont, 38 Fletcher Place, Burlington, VT 05401-1419.**
- **For the Journal of Counseling Psychology, submit manuscripts to Jo-Ida C. Hansen, PhD, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455-0344.**
- **For the Journal of Experimental Psychology: Human Perception and Performance, submit manuscripts to David A. Rosenbaum, PhD, Department of Psychology, Pennsylvania State University, 642 Moore Building, University Park, PA 16802-3104.**

Manuscript submission patterns make the precise date of completion of the 1999 volumes uncertain. Current editors, Charles R. Schuster, PhD; Clara E. Hill, PhD; and Thomas H. Carr, PhD, respectively, will receive and consider manuscripts through December 31, 1998. Should 1999 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2000 volumes.