

Overcoming Difficulties in Bayesian Reasoning: A Reply to Lewis and Keren (1999) and Mellers and McGraw (1999)

Gerd Gigerenzer and Ulrich Hoffrage
Max Planck Institute for Human Development

Bayesian reasoning can be improved by representing information in frequency formats rather than in probabilities. This thesis opens up applications in medicine, law, statistics education, and other fields. The beneficial effect is no longer in dispute, but rather its cause and its boundary conditions. C. Lewis and G. Keren (1999) argued that the effect of frequency formats is due to “joint statements” rather than to “frequency statements.” However, they overlooked the fact that our thesis is about frequency formats, not just any kind of frequency statements. We show that joint statements alone cannot account for the effect. B. A. Mellers and A. P. McGraw (1999) proposed a boundary condition under which the beneficial effect is reduced. In a reanalysis of our original data, we found this reduction for the problem they used but not for any other problem. We conclude by summarizing results indicating that teaching frequency representations fosters insight into Bayesian reasoning.

Degrees of uncertainty can be represented in various ways, including probability and frequency formats. Let us first illustrate a frequency format and how it improves Bayesian reasoning in medical experts. We asked a sample of 48 physicians with an average of 14 years of professional experience, including private practitioners, university professors, and clinic directors (Hoffrage & Gigerenzer, 1998), to make inferences about the presence of a disease given a positive result for four routinely used medical diagnostic tests. One was mammography. The relevant information (concerning a population of women aged 40 years) was presented to half of the physicians in a probability format, which can be summarized as follows: The probability of breast cancer is 1%; the probability of a positive test given breast cancer is 80%; and the probability of a positive test given no breast cancer is 10%. The question was What is the probability that a woman who tests positive actually has breast cancer? The other half of the physicians in the study received the same information in a frequency format: 10 of every 1,000 women have breast cancer; 8 of those 10 women with breast cancer will test positive; and 99 of the 990 women without breast cancer will also test positive. The question was How many of those who test positive actually have breast cancer?

When the information concerning mammography and breast cancer was presented in a probability format, only 8% of the physicians gave an estimate close to that yielded by Bayes's rule (i.e., .075). When the information was presented in a frequency format, in contrast, 46% of them arrived at the Bayesian response. This beneficial effect of the frequency format on physicians' judgments was obtained in each of the four diagnostic tasks. Thus, frequency formats help to improve Bayesian reasoning not only in

laypeople (Betsch, Biel, Eddebuttel, & Mock, 1998; Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) but in experts as well.

We consider this result exciting both in its potential for application and in its theoretical implication that reasoning is “performed” partly outside of the mind by the representation of information. There seems to be a consensus about the effect's existence. The debate has now shifted to the questions of why the effect occurs and what its boundary conditions are. These are the main issues raised by the two pairs of commentators to whom we respond here. Before we address these questions, we clarify what frequency formats are because Lewis and Keren (1999; unlike Mellers & McGraw, 1999) seem to have overlooked the fact that frequency formats are not just any kind of frequency statements.

What Are Natural Frequencies (Frequency Formats)?

A frequency format is the outcome of natural sampling rather than systematic sampling, and consists of frequencies that have not been normalized with respect to the base rates (Gigerenzer & Hoffrage, 1995, pp. 686–687). Because the term *frequency format* has been repeatedly misunderstood as frequencies of any kind, in recent publications we have adopted the (synonymous) term *natural frequencies* instead.

Natural sampling is the process of encountering instances in a population sequentially. The outcome of natural sampling is natural frequencies. Figure 1A illustrates the natural frequencies observed in a sample of 1,000 women, of which 10 have breast cancer and 990 do not.

There are two ways to arrive at frequencies that are not natural frequencies. The first is through systematic sampling, in which the base rates are fixed before any observations are made. Systematic sampling is typically used in experiments that test the effect of a treatment. Figure 1B shows an example in which 1,000 women with and 1,000 women without breast cancer were tested. Note that Figure 1B contains absolute frequencies that were not obtained through natural sampling. Unlike natural frequencies, they do not contain information about the base rates of women with and without cancer.

Gerd Gigerenzer and Ulrich Hoffrage, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development.

Correspondence concerning this article should be addressed to Gerd Gigerenzer, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Electronic mail may be sent to gigerenzer@mpib-berlin.mpg.de.

		Natural Frequencies			
		H	not H		
D	8	99			
not D	2	891			
base rate:	10	990			

A

		Normalized Frequencies			
		H	not H	H	not H
D	800	100	.8	.1	
not D	200	900	.2	.9	
	1,000	1,000	1.0	1.0	

B

C

Figure 1. A: natural frequencies. B: absolute frequencies that are not natural frequencies (obtained by systematic sampling or by normalizing natural frequencies with respect to base rates). C: relative frequencies or probabilities. *H* and *D* stand for hypothesis and data, respectively.

There is a second way to arrive at frequencies that are not natural frequencies: by normalizing natural frequencies with respect to the base rates (i.e., by setting the base rates to the same value, such as 1,000 [see Figure 1B] or 1.0 [see Figure 1C]). For instance, consider the 10 women with cancer in Figure 1A (left column), of whom 8 test positive. Normalizing this natural frequency results in 800 of 1,000 (see Figure 1B), or .8 (see Figure 1C). Normalized natural frequencies, like absolute frequencies obtained through systematic sampling, thus have the base-rate information filtered out of them.

Natural frequencies such as those in Figure 1A result from the most common form of direct observation (outside of systematic experimentation in science). Young children can count events from an early age, but do not understand fractions and other kinds of normalized counts until much later in their development (Dehaene, 1997). Examples of natural frequencies in science can be found in medical and epidemiological screening data from which base rates, hit rates, and false-positive rates are derived. The qualifier "natural" in the terms *natural sampling* and *natural frequencies* emphasizes that they are based on observations made in an ecological (rather than an experimental) setting and on raw (rather than normalized) counts of events.

The important point is that natural frequencies facilitate Bayesian computations. This is because they carry information about

base rates, whereas normalized frequencies and probabilities do not. If information is presented in normalized values, one has to multiply these by the base rates to bring the base rates "back in." Natural frequencies need not be multiplied in this way (Gigerenzer & Hoffrage, 1995).

To summarize, natural sampling yields natural frequencies, which carry information about base rates and thereby facilitate Bayesian computations. Systematic sampling and normalization do not lead to natural frequencies. Systematic sampling does not capture information about base rates because they are fixed *before* the observations are made. Normalization entails discarding the base-rate information *after* the observations are made.

With this definition of natural frequencies in mind, let us first consider Lewis and Keren's experiment and then their theoretical argument.

Normalized Frequencies Are Not Natural Frequencies

Lewis and Keren attempt to show that it is not "the use of frequencies by itself" (p. 411) that improves Bayesian reasoning. In their experiment, they compare natural frequencies (as in Figure 1A) with normalized frequencies (as in Figure 1B; they call the latter "conditional frequencies"), and report that the former generated more Bayesian responses. From this difference, they conclude that it is not the use of "frequency statements" that causes the improvement in performance, a result they believe undermines our argument about natural frequencies. They thereby overlook the point of our hypothesis, namely that Bayesian reasoning is facilitated by natural frequencies, not by just any kind of frequency statements.

In Experiment 2 (Gigerenzer & Hoffrage, 1995), we showed that relative frequencies like those in Figure 1C resulted in the same low Bayesian performance as probabilities, a result Lewis and Keren do not mention. Lewis and Keren have since shown the same result for the second type of normalized frequencies (Figure 1B).

Can Joint Statements Account for the Effect of Natural Frequencies?

We have argued that natural frequencies improve Bayesian performance because (a) they carry information about the base rates, which facilitates Bayesian computations, and (b) these computations are performed on natural numbers (Gigerenzer & Hoffrage, 1995, Results 1–7, pp. 687–689).

In contrast, Lewis and Keren claim that the beneficial effect is due not to natural frequencies but to the use of "joint statements" (i.e., statements about the joint events *D & H* and *D & not-H*) as opposed to "conditional statements." Their experiment, however, cannot distinguish between these two alternative explanations, because their joint statements were natural frequencies, and what they call conditional statements were normalized frequencies. Because natural frequencies necessarily involve joint statements, the critical test case would be joint statements that are *not* natural frequencies. A representation in terms of joint probabilities, such as $p(D \& H)$, permits such a test.

Although Lewis and Keren include joint probabilities in their Figure 1, they did not test joint probabilities. We did (Gigerenzer & Hoffrage, 1995, pp. 692–695), a fact that Lewis and Keren fail

to mention. We tested a full 2×2 factorial design that included joint probabilities, standard probabilities, and the two corresponding types of natural frequencies (Figure 2).

If the effect of frequency formats were mainly due to the use of joint statements, as Lewis and Keren claim, then a probability format with joint statements should lead to about the same performance as a frequency format. This is not the case. The probability format with joint statements yielded an average of only 28% Bayesian responses, whereas the corresponding frequency format yielded 50% (see Figure 2). The superiority of this frequency format was observed in each of the 15 problems studied. If one adds the other two cells of Figure 2 to the picture, one can see that each of the frequency formats (standard and short menu, between which Lewis and Keren do not distinguish) results in more Bayesian responses than each of the two probability formats.

To summarize, contrary to Lewis and Keren's hypothesis, joint statements alone cannot account for the beneficial effect of frequency formats.

Rare Events

In contrast to Lewis and Keren, Mellers and McGraw do not dispute that the beneficial effect of information representation that we found is due to natural frequencies, which, as they phrase it, help people to visualize "nested sets." However, they propose a condition under which the effect is no longer obtained.

Mellers and McGraw distinguish between problems that include rare as opposed to common events. They define rare events as having probabilities of .05 or lower. The boundary condition Mellers and McGraw propose is as follows: If all events are common, then the advantage of natural frequencies is reduced. If at least one of the events is rare, however, natural frequencies should have a large effect on performance.

		Menu	
		Short (joint)	Standard
Format	Frequency	50	46
	Probability	28	16

Figure 2. The 2×2 factorial design in Gigerenzer and Hoffrage (1995, Experiment 1) and the percentages of Bayesian responses in the four conditions. The standard menu displays $p(H)$, $p(D|H)$, and $p(D| \text{not-}H)$, or the corresponding frequencies, whereas the short menu displays $p(D \& H)$ and $p(D)$, or the corresponding frequencies. (Probability formats with short menus are called *joint probabilities* by the commentators.) The two probability formats are identical to the PC and PJ conditions in Lewis and Keren's (1999) Figure 1; the frequency formats, however, are not (see text). The percentages of Bayesian responses in each of the four conditions are averaged across 15 problems (including the mammography and cab problems, which the commentators examined).

Mellers and McGraw then predict that the difference between natural frequencies and probabilities will be minimized for problems with a common base rate of 15% compared with a problem with a rare base rate of 1%. Their experimental results are shown in Figure 3 (broken lines). When the base rate was 15%, the two probability formats indeed resulted in almost the same proportion of Bayesian responses as the corresponding frequency formats.

Mellers and McGraw tested their hypothesis for only one rare base rate and one common base rate. To test the generality of their finding, we reanalyzed the results of all 15 problems in our Experiment 1, which include base rates ranging from 0.005% to 80% (i.e., from very rare to very common). Figure 3 (solid lines) shows that the interaction Mellers and McGraw predict—that the difference between probabilities and natural frequencies would be smaller for problems that have only common events—was not observed. However, among the four problems with common events (base rates of 15%, 30%, 36%, and 80%), there was one (base rate of 15%) in which the proportion of Bayesian responses for the two frequency formats was almost as low as for the corresponding probability formats. This exception was the cab problem, which elicited the same effect in Mellers and McGraw's study.

This reanalysis of our data indicates that Mellers and McGraw's results are no fluke. There seems to be something special about the cab problem that suppresses or counteracts the beneficial effect of natural frequencies. Thus, there is a boundary condition on the effect of natural frequencies, but it seems to be even more specific than Mellers and McGraw state; it has to do with some feature(s) of the cab problem that reduce(s) the effect of natural frequencies. However, common events do not seem to be a boundary condition because the beneficial effect of natural frequencies was observed across a broad range of rare and common events. Instead of the predicted interaction, our reanalysis revealed a small main effect: In each of the four forms of representation tested in Experiment 1, Bayesian responses were, if anything, slightly more frequent when events were common rather than rare.

To summarize, Mellers and McGraw's results are consistent with those obtained in our Experiment 1. However, they hold for the cab problem and not for any of the other problems, some of which involve even larger base rates. The feature of the cab problem responsible for this result remains to be discovered. Thus, the boundary condition seems to be even more specific than proposed.¹

¹ Why is the overall level of performance among the participants in Mellers and McGraw's study not even half that among our University of Salzburg students (see Figure 3)? Mellers and McGraw suggest practice as an explanation: Whereas their student participants received only 1 problem each, ours worked on 30 problems in all (15 problems \times 2 representations). To test this hypothesis, we reanalyzed our data and found that the performance of our participants on the first problem was even slightly higher than that on the remaining 29 problems, lending no support to an explanation in terms of improvement by practice. Mellers and McGraw's second hypothesis, cultural differences in basic mathematical training, seems more promising to us.

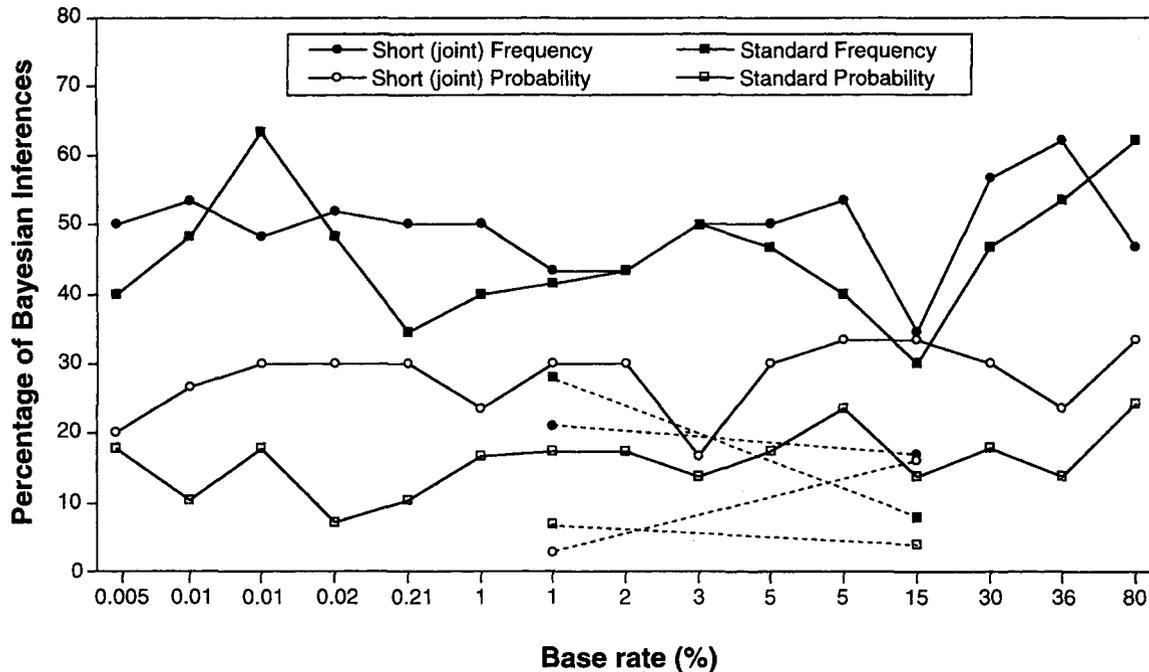


Figure 3. The 15 problems (Gigerenzer & Hoffrage, 1995, Experiment 1) ordered from rare to common base rates. Straight lines show the results of our Experiment 1; dashed lines show the results of Mellers and McGraw's (1999) study. Note that in the four problems with base rates higher than 5%, all other probabilities are also above 5%; therefore, all events are "common" in the sense of Mellers and McGraw. The percentages of Bayesian inferences for the cab problem are slightly different from those reported in Mellers and McGraw's Table 6 because the latter were based on the numerical responses alone, whereas ours were based on the numerical responses and the "write-aloud" protocols during the test and an interview after the test, as described in Gigerenzer and Hoffrage (1995, p. 692).

Multiple Representations: When Should We Use Natural Frequencies?

Lewis and Keren argue that one should represent information in terms of probabilities rather than natural frequencies because the former "reflect what we actually do in inferential science" (p. 414). We disagree with the idea that there is a single best form of information representation for all purposes. There are situations in mathematics and beyond in which probabilities are an excellent choice for representation. In experiments based on systematic sampling, for instance, no natural frequencies are obtained, and probabilities of the type $p(D/H)$ are often all that is needed, as Lewis and Keren correctly observe. (Few scientists actually use Bayesian statistics to test their hypotheses.) However, when we turn to situations in which Bayesian reasoning is important, and intuitive understanding and successful communication are crucial, then natural frequencies are a better choice. We provide two examples.

Consider a typical counselor among the 20 professional AIDS counselors we studied during actual counseling sessions with a low-risk male client (Gigerenzer, Hoffrage, & Ebert, 1998). Like all counselors, this one explained the uncertainties to the client in probabilities and percentages. For instance, he said that the prevalence of HIV among men like the client is 0.1% or slightly higher, and that the HIV test's sensitivity (the probability of a positive test given HIV) and specificity (the probability of a negative test given

no HIV) are both 99.9%. When the client asked what this would mean for the chances that he actually has HIV if he tested positive, the counselor responded that this chance is also 99.9%.

If this counselor had been trained to represent uncertainties in natural frequencies rather than in probabilities (which none of the counselors we studied had been), he would easily have understood that this answer makes no sense. Using natural frequencies, the same counselor could have explained, "Imagine 1,000 low-risk men like you taking an HIV test. One of these is infected and will test positive with practical certainty. Of the remaining 999 uninfected men, we expect that 1 will also test positive. Thus, we expect that of the 2 who test positive, only 1 is infected. Therefore, if you tested positive, your chances of really being infected would be about 1 in 2." Using natural frequencies, both counselor and client can more easily understand what a positive test means. Here, a frequency representation can prevent psychological distress based on misunderstood probabilities.

Representing information in natural frequencies can also make a difference in the courtroom, for instance, in testimony concerning DNA fingerprinting. The lawyers on O. J. Simpson's multimillion-dollar defense team are the only ones of whom we know who have taken notice of research into the power of frequency formats. They opposed the admission of conditional probabilities and likelihood ratios, which are ratios of conditional probabilities, in the testimony of DNA expert and witness for the prosecution Professor

Bruce Weir. The defense team requested that the jurors hear the information in frequencies instead, to which Judge Ito and the prosecution agreed (Koehler, 1996). The request was based on the sound argument that jurors might confuse random-match probabilities with statements about the probability that Mr. Simpson was actually the source of the samples.

These two examples illustrate real-world situations in which using natural frequencies can foster understanding (for more, see Gigerenzer, 1998).

Base Rates Vary, but Conditional Probabilities Do Not. True?

Lewis and Keren argue that one can apply the same conditional probabilities to new populations with different base rates, and that this ease of generalization favors the standard probability format. To support this argument, they claim that although women in different age groups can be expected to have different base rates of breast cancer, the “conditional rates of a positive report, given the presence or absence of breast cancer, might be expected to remain relatively stable across different populations of women. To apply them to a new group, all that would be needed is the base rate of breast cancer for the new group” (p. 414).

The problem with this argument—that base rates vary while conditional probabilities either vary little or not at all—is that base rates and conditional probabilities often vary together. For instance, the sensitivity (hit rate) of mammography is consistently lower in younger women, who also have lower base rates of breast cancer, probably because of the lower contrast between cancerous and normal glandular tissue in younger women (Fletcher, Black, Harris, Rimer, & Shapiro, 1993). Aside from specific causal factors such as the fact that time softens tissue, which in turn improves test sensitivity, there is also a more general reason for this dependency. When tests such as those for breast cancer and HIV are based on continuous rather than binary variables, one needs to define what counts as a positive result. This definition depends on the costs of the two possible errors one can make (i.e., misses and false alarms), which in turn depend on the base rates. In other words, the balance between sensitivity and specificity of a test is actually chosen dependent on the base rates.

Ignoring the fact that base rates and conditional probabilities are often dependent can lead researchers to mistake people’s sensitivity to this dependency for base-rate neglect, as demonstrated for the cab problem by Birnbaum (1983) and for the engineer-lawyer problem by Mueser, Cowan, and Mueser (1999).

To conclude, the claim that base rates and conditional probabilities are independent stands or falls with the empirical evidence. Instead of making the general assertion that base rates shift with population whereas conditional probabilities do not, we need to consult the specific evidence in each case.

Teaching Frequency Representations

One of the physicians in the study mentioned in the opening paragraph wrote to us that she now represents probabilities as natural frequencies, and no longer has “to muddle through” statistical information pretending to herself and her patients that she understands. Others remarked that natural frequencies are “more visual” and help them to see through the numbers (Hoffrage &

Gigerenzer, 1998). On the basis of these and other experiences in working with experts and laypeople, we disagree with Lewis and Keren’s unsupported claims that substituting frequency for probability representations is “mechanical” and that the “effectiveness of such attempts [to overcome cognitive biases] may be more apparent than real” (p. 414).

More systematic evidence that natural frequencies foster insight comes from studies on training in statistical reasoning. Two important measures of statistical comprehension are generalization to new problems and temporal stability after training. In testing a computer-based program for training Bayesian reasoning, Sedlmeier (1997; Sedlmeier & Gigerenzer, 1999, Experiment 1) examined both. One group of participants was taught how to represent probabilities as natural frequencies. Another group received traditional rule training, in which participants learned to insert the appropriate probabilities into Bayes’s rule. To compare the effects of representation and rule training, Sedlmeier gave participants test problems in which all statistical information was expressed in *probabilities*. The immediate generalization effect for the representation training was about twice as high as that for the rule training. Five weeks after the training, the median percentage of Bayesian responses in the rule group was down to only 15%, whereas that in the frequency representation group remained a high 90%.

Translating probabilities into natural frequencies thus proved highly effective in training Bayesian reasoning, suggesting that rather than being superficial or mechanical, natural frequencies can foster statistical insight.

Cognitive Illusions and Frequencies

Several other “cognitive illusions” can be reduced or eliminated by presenting information in frequency formats or by asking questions in terms of frequencies. We emphasize that there are different causal factors behind these effects, of which we have studied three. First, natural frequencies facilitate Bayesian computations, as in the research discussed here. Second, posing a question in terms of frequencies can clarify (or change) the reference class on which a probability judgment is based. For instance, we have argued that the “overconfidence bias” disappears with frequency questions because they switch the reference class from which the numerical judgments are derived (Gigerenzer, Hoffrage, & Kleinbölting, 1991). Third, asking a frequency question may induce the listener to interpret a question as being related to mathematical probability, whereas asking “how probable” an event is may elicit one of the legitimate nonmathematical interpretations of probability, which are often cued by other aspects of the problem. The so-called conjunction fallacy in the Linda problem is an example of an illusion that is reduced by a frequency representation for this reason (Hertwig & Gigerenzer, in press).

Although in each case a cognitive illusion is reduced or eliminated, one should not confuse the different causal factors behind the frequency effect. Also, there are more than the three just mentioned: Frequencies are easier to visualize, and they facilitate internal representations that involve discrete elements, such as in Johnson-Laird’s (1983) mental models.

Conclusions and Open Questions

Humans seem to be developmentally and evolutionarily prepared to handle natural frequencies. In contrast, many of us go through a considerable amount of mental agony to learn to think in terms of fractions, percentages, and other forms of normalized counts. An evolutionary point of view draws our attention to information representation in the environment and, specifically, to how representations change over time. This is a research heuristic of which few theorists of reasoning, judgment, and decision making take advantage. Changes in external representation, such as the historical transition from raw counts to normalized probabilities in expressing uncertainty, can suggest how numbers can be communicated most effectively to the unaided mind. Moreover, observing how the mind responds to such changes can help us to understand its design.

We hope that this reply clarifies the misunderstandings about what natural frequencies are. Contrary to Lewis and Keren's argument, the beneficial effect of natural frequencies cannot be explained by joint statements alone. Mellers and McGraw's suggested boundary condition was replicated in our experiment, but only for the cab problem.

In conclusion, we would like to mention some open questions with which we struggle. Does the beneficial effect of natural frequencies hold when there is more than one piece of data (e.g., several test results rather than one mammogram; see Krauss, Martignon, & Hoffrage, in press)? Alternatively, does the mind resort to fast and frugal heuristics as the frequency calculations become more complex (Gigerenzer, Todd, & the ABC Research Group, 1999)? If so, what do these heuristics look like? What structural properties of environments make these heuristics accurate? Mellers and McGraw's distinction between rare and common events, for instance, may point to such a structural property. Small base rates combined with large hit rates, for instance, allow for shortcuts in Bayesian reasoning with little decrement in accuracy (Gigerenzer & Hoffrage, 1995).

We invite readers to view natural frequencies as a cognitive tool with which psychologists can help experts and laypeople alike to understand the implications of statistical information. The effect of natural frequencies on Bayesian reasoning has practical consequences for education, public policy, and risk communication as well as implications for theories of judgment under uncertainty. In real-world contexts such as mammography, DNA fingerprinting, and HIV counseling, the difference in intelligibility between probabilities and natural frequencies can make the difference between despair and hope.

References

- Betsch, T., Biel, G. M., Eddelbuttel, C., & Mock, A. (1998). Natural sampling and base-rate neglect. *European Journal of Social Psychology, 28*, 269–273.
- Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96*, 85–94.
- 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.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Fletcher, S. W., Black, W., Harris, R., Rimer, B. K., & Shapiro, S. (1993). *Report of the International Workshop on Screening for Breast Cancer, February 24–25, 1993*. Bethesda, MD: National Cancer Institute.
- Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In 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, 102*, 684–704.
- Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counseling for low-risk clients. *AIDS Care, 10*, 197–211.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Hertwig, R., & Gigerenzer, G. (in press). The “conjunction” fallacy revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12*.
- Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine, 73*, 538–540.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review, 67*, 859–886.
- Krauss, S., Martignon, L., & Hoffrage, U. (in press). Simplifying Bayesian inference. In L. Magnani, N. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery*. New York: Plenum.
- Lewis, C., & Keren, G. (1999). On the difficulties underlying Bayesian reasoning: A comment on Gigerenzer and Hoffrage. *Psychological Review, 106*, 411–416.
- Mellers, B. A., & McGraw, A. P. (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage. *Psychological Review, 106*, 417–424.
- Mueser, P. R., Cowan, N., & Mueser, K. T. (1999). A generalized signal detection model to predict rational variation in base rate use. *Cognition, 69*, 267–312.
- Sedlmeier, P. (1997). BasicBayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments, & Computers, 29*, 328–336.
- Sedlmeier, P., & Gigerenzer, G. (1999). *Teaching Bayesian reasoning in less than two hours*. Manuscript submitted for publication.

Received May 4, 1998

Revision received December 7, 1998

Accepted December 10, 1998 ■