In R. Bibace, J. D. Laird, K. L. Noller & J. Valsiner (Eds.). (2005). *Science and medicine in dialogue* (pp. 83–98). Westport, CT: Praeger.

Chapter 5

UNDERSTANDING THE RESULTS OF MEDICAL TESTS: WHY THE REPRESENTATION OF STATISTICAL INFORMATION MATTERS

Ulrich Hoffrage, Stephanie Kurzenhäuser, and Gerd Gigerenzer

Women are generally informed that mammography screening reduces the risk of dying from breast cancer by 25 percent. Does that mean that for every 100 women who participate in screening, 25 lives will be saved? Although many people believe this to be the case, it is incorrect. The percentage means that for 1,000 women who participate in screening, 3 will die from breast cancer within 10 years, whereas for 1,000 women who do not participate, 4 will die. The difference between 4 and 3 is the 25 percent "relative risk reduction." Expressed as an "absolute risk reduction," the benefit is 1 in 1,000.

The topic of this chapter is the representation of information on medical risks. As the case of mammography screening illustrates, the same information can be presented in various ways. The choice among alternative representations can influence patients' hopes and fears, risks and choices, and ultimately their behavior. For example, women were most likely to accept screening for cancer when the benefits of screening were presented as a relative risk reduction, less likely to do so when the absolute risk reduction was used, and least likely when the benefits were presented in terms of the numbers of women that need to be screened in order to save one life (Sarfati, Howden-Chapman, Woodward, & Salmond, 1998). This observation leaves us with a dilemma. According to Sarfati et al. (1998), health professionals have to make a choice. In order to enhance participation rates, they can either frame the benefits of screening in the most positive light, or they can present the information to reduce framing effects—for example, by expressing the benefits in a variety of forms. The authors contend that there may be a tension between these approaches. While the former is arguably manipulative, the latter may enhance informed choice but reduce participation rates in screening programs.

In our view, high participation rates should not be an ideal per se. Instead, each woman should be helped to understand the pros and cons of screening, to clarify her own values, and to consider the decision that would be best for her. Informed consent involves more than signing a form. In the present chapter, we assume that the patient and physician share the same goal, namely, to reach such an informed decision, based on the patient's understanding of the benefits and risks of a treatment or the chances that a particular diagnosis is right or wrong. There are two necessary steps toward this ideal. First, physicians themselves need to understand the statistical information and its implications, and second, they need to learn how to communicate this information to the patients. This double requirement contrasts sharply with the fact that physicians are rarely trained in risk communication. The lack of training may explain why previous research observed that a majority of physicians do not use relevant statistical information properly. Casscells, Schoenberger, and Grayboys (1978), for example, asked 60 house officers, students, and physicians at Harvard Medical School to estimate the probability of an unnamed disease, given the following information:

> If a test to detect a disease whose prevalence is 1/1,000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? (p. 999)

The estimates varied wildly, from the most frequent estimate, 95 percent (27 out of 60), to 2 percent (11 out of 60). The value of 2 percent is obtained by inserting the problem information into Bayes's rule—assuming that the sensitivity of the test, which is not specified in the problem, is approximately 100 percent. Casscells et al. (1978) concluded that "in this group of students and physicians, formal decision analysis was almost entirely unknown and even common-sense reasoning about the interpretation of laboratory data was uncommon" (p. 1000).

In an article on probabilistic reasoning about mammography, Eddy (1982) reported an informal study in which he asked several physicians to estimate the probability of breast cancer given a base rate (*prevalence*) of 1 percent, a hit rate (*sensitivity*, here the proportion of positive mammograms among women with breast cancer) of 79 percent, and a false-positive rate (proportion of positive mammograms among women without breast cancer; the complement of the *specificity*) of 9.6 percent. He reported that 95 out of 100 physicians estimated the posterior probability of breast cancer given a positive mammogram (the *positive predictive value*) to be between 70 percent and 80 percent, whereas Bayes's rule results in a value one order of magnitude smaller, namely, 7.7 percent. Eddy proposed that the majority of physicians confused the sensitivity of the test with the positive predictive value. Evidence of this confusion can also be found in medical textbooks and journal articles (Eddy, 1982) as well as in statistical textbooks (Gigerenzer, 1993).

In 1986, Windeler and Köbberling reported responses to a questionnaire they mailed to family physicians, surgeons, internists, and gynecologists in Germany. Only 13 of the 50 respondents realized that an increase in the prevalence of a disease implies an increase in the positive predictive value. The authors concluded with a puzzling observation. Although intuitive judgment of probabilities is part of every diagnostic and treatment decision, the physicians in their study were unaccustomed to estimating quantitative probabilities. Given these demonstrations that many physicians' reasoning does not follow the laws of probability (Abernathy & Hamm, 1995; Dawes, 1988; Dowie & Elstein, 1988), what can be done to improve diagnostic inferences? In the remainder of this chapter we propose an easy way to help physicians and patients understand statistical information and its consequences, and we report empirical evidence from three studies demonstrating the benefits of the proposed method. We conclude with a discussion of the impact of this research on medical education, AIDS counseling, and DNA fingerprinting.

NATURAL FREQUENCIES HELP IN MAKING DIAGNOSTIC INFERENCES

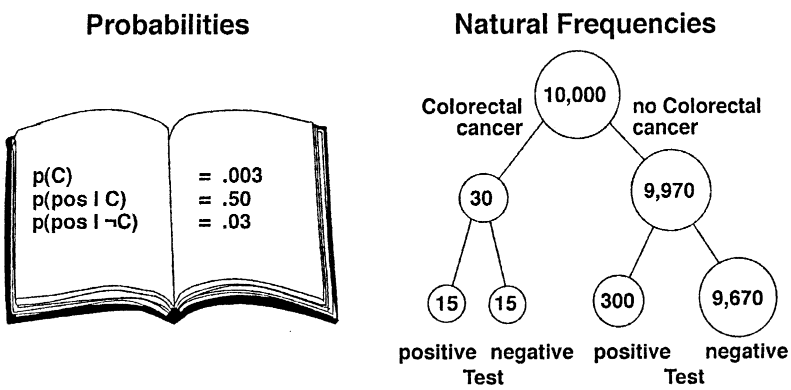
Each of the three studies, summarized above, presented numerical information in the form of probabilities and percentages. The same holds for other studies in which the conclusion was that physicians (Berwick, Fineberg, & Weinstein, 1981; Politser, 1984) and lay persons (Koehler, 1996a) have great difficulty in making diagnostic inferences from statistical information. Whether the information is presented as probabilities, percentages, absolute frequencies, or another form is irrelevant from a mathematical viewpoint. However, they are not equivalent from a psychological viewpoint, which is the key to our argument.

We argue that a specific class of representations, which we call *natural frequencies,* help lay persons and experts make inferences the Bayesian way. We illustrate the difference between probabilities and natural frequencies with the diagnostic problem of inferring the presence of colorectal cancer (C) from a positive result in the hemoccult test (pos), a standard diagnostic test. In terms of probabilities, the relevant information is a base rate for colorectal cancer [p(C)] of 0.3 percent, a sensitivity [p(pos|C)] of 50 percent, and a false-positive rate [p(pos|¬C)] of 3 percent. In natural frequencies, the same information would read, "Thirty out of every 10,000 persons have colorectal cancer. Of these 30 with cancer, 15 will have a positive hemoccult test. Of the remaining 9,970 people *without* colorectal cancer, 300 will have a positive hemoccult test." Natural frequencies are absolute frequencies, as they result from sequentially encoding and aggregating observations from a population (or a representative sample). Natural frequencies have not been normalized with respect to the base rates of disease and nondisease (Gigerenzer & Hoffrage, 1995, 1999). Natural frequencies should be distinguished from probabilities, percentages, relative frequencies, and other representations where the underlying natural frequencies have been normalized with respect to these base rates. For example, the following representation of the colorectal cancer problem is not in terms of natural frequencies (Gigerenzer & Hoffrage, 1995), because the frequencies have been normalized with respect to the base rates: a base rate of 30 in 10,000, a sensitivity of 5,000 in 10,000, and a false-positive rate of 300 in 10,000.

Why should natural frequencies facilitate diagnostic inferences? There are two related arguments. The first is computational. Bayesian computations are simpler when the information is represented in natural frequencies rather than in probabilities, percentages, or relative frequencies (Christensen-Szalanski & Bushyhead, 1981; Kleiter, 1994). For example, when the information concerning colorectal cancer is represented in probabilities, applying a cognitive algorithm to compute the positive predictive value, that is, the Bayesian posterior probability, amounts to performing the computation shown in the left side of Figure 5.1. The result is 0.048. This equation is Bayes's rule for binary hypotheses (here, C and ¬C) and data (here, positive and negative test result). The rule is named after Thomas Bayes (1702–1761), who is credited with solving the problem of how to make an inference from data to hypothesis (Stigler, 1983). As can be seen from the right side of Figure 5.1, the computations are much simpler when the information is presented in natural frequencies. The equation in the right box is Bayes's rule for natural frequencies, where [C&pos] is the number of cases with cancer and a positive test, and [¬C&pos] is the number of cases without cancer but with a positive test.

The second argument supplements the first. Minds appear to be tuned to make inferences from natural frequencies rather than from probabilities and percentages. This argument is consistent with developmental studies indicating the primacy of reasoning with discrete numbers over fractions, and studies of adult humans and animals indicating the ability to monitor frequency information in natural environments in fairly accurate and automatic ways (Gallistel & Gelman, 1992; Jonides & Jones, 1992; Real, 1991; Sedlmeier, Hertwig, & Gigerenzer, 1998). For most of their existence, humans have made inferences from information encoded sequentially through direct experience, and natural frequencies are

Figure 5.1
Why natural frequencies facilitate the computation of the probability p(C|pos) of cancer given a positive test (a form of Bayesian reasoning). The symbols "C" and "¬C" stand for colorectal cancer and no colorectal cancer, respectively, and "pos" and "neg" stand for a positive and negative test result, respectively. One can see that Bayes's rule for probabilities (left side) involves more calculation than that for natural frequencies (right side).



$$p(C \mid pos) = \frac{p(C)\,p(pos \mid C)}{p(C)\,p(pos \mid C) + p(\neg C)\,p(pos \mid \neg C)} = \frac{(.003)\,(.50)}{(.003)\,(.50) + (.997)\,(.03)} = .047$$

$$p(C \mid pos) = \frac{C\&pos}{C\&pos + \neg C\&pos} = \frac{15}{15 + 300} = .047$$

the final tally of such a process. Mathematical probability emerged in the mid-seventeenth century (Daston, 1988), and not until after the French Revolution did percentages appear to have become commonly used, mainly for taxes and interest, and only very recently for risk and uncertainty (Gigerenzer, Swijtink, Porter, Daston, Beatty, & Krüger, 1989). Minds might have evolved to deal with natural frequencies rather than with probabilities.

We tested whether natural frequencies improve Bayesian inference in lay persons, medical students, and physicians.

DO NATURAL FREQUENCIES IMPROVE LAYPERSONS' REASONING?

We first tested students in various fields at the University of Salzburg (Gigerenzer & Hoffrage, 1995). We used 15 problems, including Eddy's mammography problem and Tversky and Kahneman's (1982) cab problem. When the information was presented in natural frequencies rather than in probabilities, the proportion of Bayesian responses increased for each of the 15 problems. The average proportions of Bayesian responses were 16 percent for probabilities and 46 percent for natural frequencies (Gigerenzer & Hoffrage, 1995). Similarly, Cosmides and Tooby (1996) showed that natural frequencies improve Bayesian inferences in the Casscells et al. (1978) problem as well. This hypothetical medical problem is numerically simpler (the hit rate is assumed to be 100%) than the problems in the Gigerenzer and Hoffrage (1995) study, and Cosmides and Tooby reported that 76 percent of the answers were Bayesian (see also Christensen-Szalanski & Beach, 1982). These results lead us to conclude that natural frequencies improve Bayesian reasoning without instruction, at least in laypersons.

But would medical experts also profit from natural frequencies, and do they use them in communicating risks to their clients? The following two studies with medical students and experienced physicians provide an answer to the first question; a study with AIDS counselors addresses the second question.

DO NATURAL FREQUENCIES IMPROVE MEDICAL STUDENTS' DIAGNOSTIC INFERENCES?

We chose four realistic diagnostic tasks and constructed two versions of each. In one, the information was presented in probabilities, and in the other, the information was presented in natural frequencies. The four diagnostic tasks were to infer (a) the presence of colorectal cancer from a positive hemoccult test, (b) the presence of breast cancer from a positive mammogram, (c) the presence of phenylketonuria from a positive Guthrie test, and (d) the presence of ankylosing spondylitis from a positive HL-antigen-B27 (HLA-B27) test. The information on prevalence, sensitivity, and false positives was taken from the literature (Hoffrage & Gigerenzer, 2004; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000).

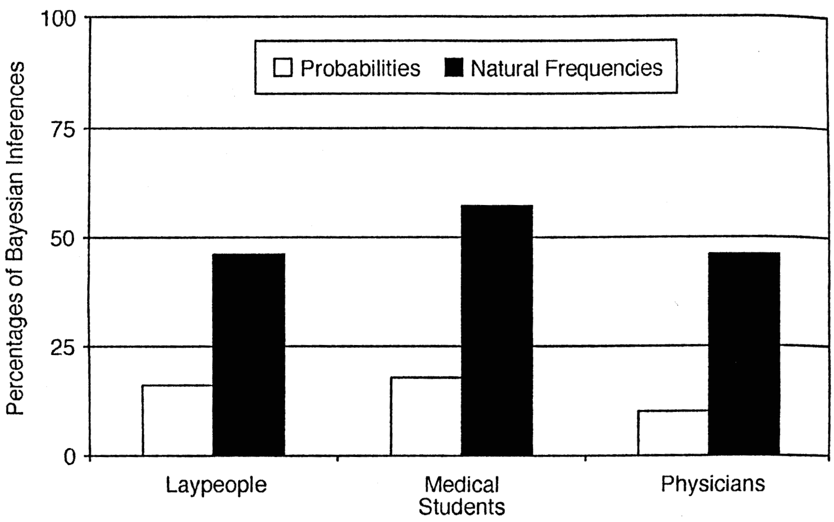
Participants were 87 medical students, most of whom had already passed a course in biostatistics and were, on average, in their fifth year, and 9 first-year interns. Fifty-four studied in Berlin, and 42 in Heidelberg; 52 were female, and 44 were male. The average age was 25 years. Participants were paid a participation fee of 15 DM (approx. $7.50). They worked on the questionnaire at their own pace and in small groups of three to six. The experimenter asked them to make notes, calculations, or drawings, so that we could reconstruct their reasoning. Interviews were performed after the participants completed their questionnaire.

When a participant's estimate was within plus or minus 5 percent of the Bayesian estimate, and the notes and interview indicated that the estimate was arrived at by Bayesian reasoning (Gigerenzer & Hoffrage, 1995) rather than by guessing or other means, we then classified the response as a "Bayesian inference." For each task, the percentages of Bayesian inferences were higher when the information was presented in natural frequencies rather than in probabilities. Across all participants and across all four problems, the percentage of Bayesian inferences was 18 percent when the information was given in probabilities and 57 percent when it was given in natural frequencies (Figure 5.2). Moreover, if we consider only the estimates that were not classified as Bayesian inferences, the absolute deviation from the Bayesian answer was 21 percent lower for the frequency problems than for the probability problems (42%) (Hoffrage & Gigerenzer, 2004). In conclusion, medical students encountered problems similar to those of laypersons when the information was in probabilities, but their reasoning improved more than laypersons when frequency representations were used.

DO NATURAL FREQUENCIES IMPROVE PHYSICIANS' DIAGNOSTIC INFERENCES?

Would these findings generalize to experienced physicians who treat real patients? Forty-eight physicians participated in the following study (Gigerenzer, 1996; Hoffrage & Gigerenzer, 1998). They had practiced

Figure 5.2
Effect of information representation (probabilities versus natural frequencies) on statistical reasoning in laypeople, medical students, and physicians, based on 15 Bayesian inference tasks for laypeople and 4 each for medical students and physicians.



Source: Data from Gigerenzer & Hoffrage, 1995; Hoffrage et al., 2000; Hoffrage & Gigerenzer, 1998.

for an average of 14 years (1 month to 32 years) and had a mean age of 42 years (26 to 59). They worked either in Munich or Düsseldorf. Eighteen were female, and 30 were male. Eighteen worked in university hospitals, 16 in private or public hospitals, and 14 in private practice. The sample included internists, gynecologists, dermatologists, and radiologists, among others. The physicians' status ranged from directors of clinics to physicians commencing their careers.

The interviewer first informed the physician about our interest in studying diagnostic inference and established a relaxed personal rapport. Each physician was then given the same four diagnostic tasks as in the previous study. Each problem was printed on a sheet of paper, and the interviewer asked the physician to make notes, calculations, or drawings so that we could later reconstruct their reasoning. After the physician completed the four tasks, if it could not be discerned how the estimate was made in each task, the physician was asked for clarification. In two diagnostic tasks, the information was presented in probabilities, and in the other two, in natural frequencies. We systematically varied which tasks were in which format and which format was presented first with the constraint that the first two tasks had the same format. To classify a strategy as Bayesian, we used the same criteria as in the previous study.

These physicians reasoned the Bayesian way more often when the information was communicated in natural frequencies than when it was communicated in probabilities. The effect varied among problems, but even in the problem showing the weakest effect (phenylketonuria), the proportion of Bayesian answers was twice as large. For the two cancer problems, natural frequencies increased Bayesian inferences by more than a factor of five as compared to probabilities. Across all problems, the physicians gave the Bayesian answer in only 10 percent of the cases that used probabilities. When natural frequencies were used, this value increased to 46 percent (Figure 5.2).

With probabilities, physicians spent an average of 25 percent more time solving the diagnostic problems than with natural frequencies. Moreover, physicians commented that they were nervous, tense, and uncertain more often when working with probabilities than when working with natural frequencies. They also stated that they were less skeptical of the relevance of statistical information when it was in natural frequencies. Physicians were conscious of their better and faster performance with natural frequencies. We asked the physicians how often they took statistical information into account when they interpreted the results of diagnostic tests. Twenty-six answered "very seldom" or "never," 15 answered "once in a while," 5 said "frequently," and none answered "always."

Their comments suggested two reasons why they used statistical information rather infrequently: the patient's uniqueness and the physician's innumeracy. The first reason can be illustrated by a comment from one of three physicians who refused to participate in the study and did not contribute to the data. This physician explained, "I can't do much with numbers. I am an intuitive being. I treat my patients in a holistic manner and don't use statistics." Similarly, a university professor who seemed agitated and affronted by the test and refused to give numerical estimates remarked, "This is not the way to treat patients. I throw all these journals [with statistical information] away immediately. One can't make a diagnosis on such a basis. Statistical information is one big lie."

The second reason for physicans' reluctance to use statistical information is related to the first. Several physicians perceived themselves as mathematically illiterate, or suffering from a cognitive disease known as "innumeracy" (Paulos, 1988). Six physicians explicitly remarked on their inability to deal with numbers, stating, for example, "But this is mathematics. I can't do that. I'm too stupid for this." With natural frequencies, however, these same physicians spontaneously reasoned statistically as often as their peers who did not complain of innumeracy.

TEACHING BAYESIAN REASONING

Only five of the 48 physicians from the last study stated that they had heard of Bayes's rule. If natural frequencies can foster Bayesian reasoning without instruction, it is straightforward to also use them in statistical education. Statistical information in medical textbooks, newspapers, and other media is most often displayed in a probability or percentage format. Perhaps training should enable participants to translate probabilities into natural frequencies.

Sedlmeier and Gigerenzer (2001) and Sedlmeier (1997) were the first to apply this idea to teaching. They designed a two-hour computerized tutorial where participants could learn to solve Bayesian tasks by translating probabilities into natural frequencies (Figure 5.1, right side). For comparison, participants in another group received the traditional training that teaches how to insert probabilities into Bayes's rule (Figure 5.1, left side). In the group who received the representation training, the proportion of Bayesian inferences in a test taken immediately after the training was substantially higher than in the group who received the traditional rule training.

But how quickly did students forget what they had learned? In several experiments, the students were re-tested between 1 week and 3 months later. The students who had gone through the rule training showed the typical forgetting effect. For example, in one study, performance was down to 20 percent after five weeks. In contrast, when students had learned frequency representations, their performance remained consistently at the level they had achieved immediately after training, which was a median of 90 percent Bayesian responses.

While the use of flexible computer-based tutorial systems has clear advantages, the range of possible applications is still limited. In German universities where traditional instruction in front of a classroom with a blackboard and an overhead projector is customary, the use of interactive tutorial programs is still the exception. Would the representation training approach still be successful when applied in this type of instructional setting? To answer this question, Kurzenhäuser and Hoffrage (2002) developed a one-hour classroom tutorial on Bayesian reasoning, based on the representation training approach, and tested it in a human genetics course for medical students. Participants were 208 medical students in an obligatory all-day course at the Free University of Berlin. To evaluate the relative effectiveness of the new approach in a classroom setting, we also included a traditional rule training condition; 109 participants received the representation training, and 99 participants received the rule training. The two approaches were evaluated two months later by testing the students' ability to solve a Bayesian inference task with information represented as probabilities. While both approaches improved performance compared to pre-test results, almost three times as many students were able to profit from the representation training as opposed to the rule training (47% and 16%, respectively).

AIDS COUNSELING FOR LOW-RISK CLIENTS

An important application of Bayesian reasoning is in AIDS counseling for low-risk clients. In Germany, the prevalence of HIV in heterosexual men who are in no known risk group is approximately 0.01 percent, the specificity of the HIV test is approximately 99.99 percent, and the sensitivity is approximately 99.9 percent. If a counselor communicates these numbers, the client will most likely not be able to work out his chances of having the virus if he tests positive. Most seem to assume that a positive test means infection. For example, in the early days of blood screening in Florida, after 22 blood donors were told they were HIV positive, seven committed suicide (Stine, 1996).

How do AIDS counselors explain what a positive test means to their clients? We studied AIDS counselors in German public health centers (Gigerenzer, Hoffrage, & Ebert, 1998). One of us visited 20 centers as a client to take 20 HIV tests and make use of the mandatory pre-test counseling. The counselor was asked the relevant questions concerning the prevalence of an HIV infection, the sensitivity and specificity of the test, and what the chances were that the client actually has the virus if the test was positive. Not one counselor communicated the risks to the client in natural frequencies. Instead, they all used probabilities and percentages, and in the majority of the counseling sessions, the information was either internally inconsistent or incorrect. For example, one counselor estimated the base rate at approximately 0.1 percent and the sensitivity and specificity at 99.9 percent, and concluded that the client's chance of having the virus if he tested positive is also 99.9 percent. In fact, 15 out of 20 counselors told their low-risk client that it is 99.9 percent or 100 percent certain that he has HIV if he tests positive.

If a counselor, however, communicates the information in natural frequencies, insight is more likely: "Think of 10,000 heterosexual men like yourself being tested. We expect that one has the virus, and this one will test positive. Of the remaining 9,999 uninfected men, one will also test positive. Thus, we expect that for every two men in this risk group who test positive, only one has HIV, or 50 percent."

In real-world contexts such as AIDS counseling, the difference between natural frequencies and probabilities can make the difference between hope and despair.

DNA FINGERPRINTING

The relevance of natural frequencies is not limited to medical diagnosis. As Koehler's work (e.g., 1996b) demonstrates, the difficulty in drawing inferences from probabilities also holds for DNA experts, judges, and prosecutors. Nevertheless, in criminal and paternity cases, the general practice in court is to present information in terms of probabilities or likelihood ratios, with the consequence that jurors, judges, and sometimes the experts themselves are confused and misinterpret the evidence. In a recent study, Hoffrage et al. (2000) demonstrated that both law students and jurists profit from natural frequencies. The percentage of Bayesian inferences rose from 3 percent to 45 percent when the format of the information concerning DNA fingerprinting changed from probabilities to natural frequencies. Possibly even more important, the participants who had seen the information in terms of probabilities had a higher conviction rate than those participants who had been given the same information in terms of natural frequencies (Hertwig & Hoffrage, 2002; Lindsey, Hertwig, & Gigerenzer, 2003).

CONCLUSIONS

Statistical reasoning is indispensable in a modern, technological democracy, similar to the ability to read and write (Gigerenzer, 2002). The last few decades have witnessed much debate on whether minds are equipped with the right or wrong rules for making judgments under conditions of uncertainty. However, the ability to draw inferences from statistical information depends not only on cognitive strategies but also on the format in which the numerical information is communicated. Insight can come from outside. External representation can perform part of the reasoning process. In our studies, natural frequencies improved both laypersons' and experts' statistical reasoning, with and without explicit teaching.

Basic research on reasoning can produce simple and powerful methods of communicating risks that can be of help in various public domains.

ACKNOWLEDGMENTS

We thank Rona Unrau for editing the manuscript and the German Research Foundation (DFG) for their financial support (Ho1847/1–2).

REFERENCES

Abernathy, C. M., & Hamm, R. M. (1995). *Surgical intuition: What it is and how to get it.* Philadelphia: Hanley & Belfus.

Berwick, D. M., Fineberg, H. V., & Weinstein, M. C. (1981). When doctors meet numbers. *American Journal of Medicine, 71,* 991–998.

Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299,* 999–1000.

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

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

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.

Daston, L. J. (1988). *Classical probability in the Enlightenment.* Princeton, NJ: Princeton University Press.

Dawes, R. M. (1988). *Rational choice in an uncertain world.* San Diego, CA: Harcourt, Brace, Jovanovich.

Dowie, J., & Elstein, A. (1988). *Professional judgment: A reader in clinical decision making.* Cambridge, England: Cambridge University Press.

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, England: Cambridge University Press.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43–74.

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 313–339). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Journal of Medical Decision Making, 16,* 273–280.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you.* New York: Simon & Schuster.

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

Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis & Keren and Mellers & McGraw. *Psychological Review, 104,* 425–430.

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

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

Hertwig, R., & Hoffrage, U. (2002). Technology needs psychology: How natural frequencies foster insight in medical and legal experts. In P. Sedlmeier & T. Betsch (Eds.), *Etc. Frequency processing and cognition* (pp. 285–302). New York: Oxford University Press.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine, 73,* 538–540.

Hoffrage, U., & Gigerenzer, G. (2004). How to improve the diagnostic inferences of medical experts. In E. Kurz-Milcke & G. Gigerenzer (Eds.), *Experts in science and society* (pp. 249–268). New York: Kluwer Academic/Plenum Publishers.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290,* 2261–2262.

Jonides, J., & Jones, C. M. (1992). Direct coding for frequency of occurrence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 368–378.

Kleiter, G. D. (1994). Natural sampling: Rationality without base rates. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375–388). New York: Springer.

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

Koehler, J. J. (1996b). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review, 67,* 859–886.

Kurzenhäuser, S., & Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher, 24,* 531–536.

Lindsey, S., Hertwig, R., & Gigerenzer, G. (2003). Communicating statistical DNA evidence. *Jurimetrics*, 43, 147–163.

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences.* New York: Vintage Books.

Politser, P. E. (1984). Explanations of statistical concepts: Can they penetrate the haze of Bayes? *Methods of Information in Medicine, 23,* 99–108.

Real, L. A. (1991). Animal choice behavior and the evolution of cognitive architecture. *Science, 253,* 980–986.

Sarfati, D., Howden-Chapman, P., Woodward, A., & Salmond, C. (1998). Does the frame affect the picture? A study into how attitudes to screening for cancer are affected by the way benefits are expressed. *Journal of Medical Screening, 5,* 137–140.

Sedlmeier, P. (1997). BasicBayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments, & Computers, 29,* 328–336.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130,* 380–400.

Sedlmeier, P., Hertwig, R., & Gigerenzer, G. (1998). Are judgments of the positional frequencies of letters systematically biased due to availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 754–770.

Stigler, S. M. (1983). Who discovered Bayes's theorem? *American Statistician, 37,* 296–325.

Stine, G. J. (1996). *Acquired immune deficiency syndrome. Biological, medical, social and legal issues.* Englewood Cliffs, NJ: Prentice Hall.

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.

Windeler, J., & Köbberling, J. (1986). Empirische Untersuchung zur Einschätzung diagnostischer Verfahren am Beispiel des Haemoccult-Tests [An empirical study of the judgments about diagnostic procedures using the example of the hemoccult test]. *Klinische Wochenschrift, 64,* 1106–1112.

EDITORIAL COMMENTARY

As this chapter clearly shows, many physicians do not understand the meaning of probabilities. The natural frequencies presentation should make sense to these clinicians. However, many patients would still find this concept confusing. Over the past thirty years I have had to tell patients what their findings mean on thousands of occasions. Because of the type of practice I have, I need to explain that they are at an increased risk for cancer, and then must find some way to quantify that risk. For some—a minority of patients—probabilities work well. For others, the natural frequencies are a meaningful method of expression. However, there remains a large fraction for which an analogy may best allow the patient to choose a course of action. For example, when a patient is overly concerned about a minor Pap smear abnormality, I will often tell her, "You are at greater risk driving home from the office than from this abnormality." On the other end of the risk spectrum, it is sometimes necessary to draw a very gloomy picture. On rare occasions I've told a patient that "You won't live to see your children graduate from high school if you don't. . . . " It is always a challenge when it is necessary to explain to a patient the risk of a certain disease or diagnosis, and the astute clinician varies the method of explanation based on a thorough knowledge of the patient.

—Kenneth L. Noller