Ecological Intelligence

An Adaptation for Frequencies

Gerd Gigerenzer

printed from:

The Evolution of Mind

Edited by

Denise Dellarosa Cummins and Colin Allen

Ecological Intelligence

An Adaptation for Frequencies

GERD GIGERENZER

When I left a restaurant in a charming town in Tuscany one night, I looked for my yellow-green rented Renault 4 in the parking lot. There was none. Instead, there was a blue Renault 4 sitting in the lot, the same model, but the wrong color. I still feel my fingers hesitating to put my key into the lock of this car, but the lock opened. I drove the car home. When I looked out the window the next morning, there was a yellow-green Renault 4 standing in bright sunlight outside. What had happened? My color-constancy system did not work with the artificial light at the parking lot. Color constancy, an impressive adaptation of the human perceptual system, normally allows us to see the same color under changing illuminations, under the bluish light of day as well as the reddish light of the setting sun. Color constancy, however, fails under certain artificial lights, such as sodium or mercury vapor lamps, which were not present in the environment when mammals evolved (Shepard, 1992).

Human color vision is adapted to the spectral properties of natural sunlight. More generally, our perceptual system has been shaped by the environment in which our ancestors evolved, the environment often referred to as the “environment of evolutionary adaptiveness,” or EEA (Tooby & Cosmides, 1992). Similarly, human morphology, physiology, and the nervous and immune systems show exquisite adaptations. The tubular form of the bones maximizes strength and flexibility while minimizing weight; bones are, pound for pound, stronger than solid steel bars, and the best man-made heart valves cannot yet match the way natural valves open and close (Nesse
& Williams, 1995). Like color constancy, however, these systems can be fooled and may break down when stable, long-term properties of the environment to which they were adapted change.

In this chapter, I propose that human reasoning algorithms are, like those of color constancy, designed for information that comes in a format that was present in the EEA. I will focus on a class of inductive reasoning processes technically known as Bayesian inference, specifically a simple version thereof in which an organism infers from one or a few indicators which of two events is true.

Bayesian Inference

David Eddy (1982) asked physicians to estimate the probability that a woman has breast cancer given that she has a positive mammogram on the basis of the following information:

The probability that a patient has breast cancer is 1% (the physician's prior probability).

If the patient has breast cancer, the probability that the radiologist will correctly diagnose it is 79% (sensitivity or hit rate).

If the patient has a benign lesion (no breast cancer), the probability that the radiologist will incorrectly diagnose it as cancer is 9.6% (false positive rate).

QUESTION: What is the probability that a patient with a positive mammogram actually has breast cancer?

Eddy reported that 95 out of 100 physicians estimated the probability of breast cancer after a positive mammogram to be about 75 percent. The inference from an observation (positive test) to a disease, or more generally, from data D to a hypothesis H, is often referred to as “Bayesian inference,” because it can be modeled by Bayes’s rule:

\[
p(H|D) = \frac{p(H)p(D|H)}{p(H)p(D|H) + p(\bar{H})p(D|\bar{H})} = \frac{(.01)(.79)}{(.01)(.79) + (.99)(.096)} = .077
\]  

(1)

Equation 1 shows how the probability \( p(H|D) \) that the woman has breast cancer (H) after a positive mammogram (D) is computed from the prior probability \( p(H) \) that the patient has breast cancer, the sensitivity \( p(D|H) \), and the false positive rate \( p(D|\bar{H}) \) of the mammography test. The probability \( p(H|D) \) is called the “posterior probability.” The symbol \( \bar{H} \) stands for “the patient does not have breast cancer.” Equation 1 is Bayes’s rule for binary hypotheses and data. The rule is named after Thomas Bayes (1702–1761), an English dissenting minister, to whom this solution of the problem of how to make an inference from data to hypothesis (the so-called inverse problem; see Daston, 1988) is attributed.¹ The important point is that Equation 1 results in a probability of 7.7%, not 75% as estimated by the majority of physicians. In other words, the probability that the woman has breast cancer is one order of magnitude smaller than estimated.
This result, together with an avalanche of studies reporting that lay
people's reasoning does not follow Bayes's rule either, has (mis-)led many
to believe that Homo Sapiens would be inept to reason the Bayesian way.
Listen to some influential voices: "In his evaluation of evidence, man is
apparently not a conservative Bayesian: he is not a Bayesian at all" (Kahneman
& Tversky, 1972, p. 450). "Tversky and Kahneman argue, correctly,
I think, that our minds are not built (for whatever reason) to work by the
rules of probability" (Gould, 1992, p. 469). The literature of the last
twenty-five years has reiterated again and again the message that people
are bad reasoners, neglect base rates most of the time, neglect false positive
rates, and are unable to integrate base rate, hit rate, and false positive rate
the Bayesian way (for a recent review see Koehler, 1996). Probability prob-
lems such as the mammography problem have become the stock-in-trade
of textbooks, lectures, and party entertainment. It is guaranteed fun to point
out how dumb others are. And aren't they? There seem to be many cus-
tomers eager to buy the message of "inevitable illusions" wired into our
brain (Piattelli-Palmarini, 1994).

Ecological Bayesian Inference: An Adaptation for Frequencies

Back to color constancy. If a human visual system enters an environment
illuminated by sodium vapor lamps, its color constancy algorithms will fail.
This does not mean, however, that human minds are not built to work by
color-constancy algorithms. Similarly, if a human reasoning system enters
an environment in which statistical information is formatted differently from
that encountered in the environment in which humans evolved, the rea-
soning algorithms may fail. But this does not imply that human minds are
not built to reason the Bayesian way. The issue is not whether nature has
equipped our minds with good or with bad statistical software, as the "opti-
mists" versus "pessimists" discussion about human rationality suggests
(Jungermann, 1983). The issue I address here is the adaptation of mental
algorithms to their environment. By "mental algorithms," I mean induc-
tion mechanisms that perform classification, estimation, or other forms of
uncertain inferences, such as deciding what color an object is, or inferring
whether a person has a disease.

For which information formats have mental algorithms been designed?
What matters for an algorithm that makes inductive inferences is the for-
mat of numerical information. Eddy presented information (about the preva-
ience of breast cancer, the sensitivity, and the false positive rate of the test)
in terms of probabilities and percentages, just as most experimenters did
who found humans making irrational judgments. What was the format of
the numerical information humans encountered during their evolution? We
know too little about these environments, for instance, about the histori-
ically normal conditions of childbirth, or how strong a factor religious do-
ctrines were, and most likely, these varied considerably between societies.
But concerning the format of numerical information, I believe we can be
as certain as we ever can be—probabilities and percentages were not the way organisms encountered information. Probabilities and percentages are quite recent forms of representations of uncertainty. Mathematical probability emerged in the mid-seventeenth century (Hacking, 1975), and the concept of probability itself did not gain prominence over the primitive notion of “expectation” before the mid-eighteenth century (Daston, 1988). Percentages became common notations only during the nineteenth century, after the metric system was introduced during the French Revolution (mainly, though, for interest and taxes rather than for representing uncertainty). Only in the second half of the twentieth century did probabilities and percentages become entrenched in the everyday language of Western countries as representations of uncertainty (Gigerenzer et al., 1989). To summarize, probabilities and percentages took millennia of literacy and numeracy to evolve as a format to represent degrees of uncertainty. In what format did humans acquire numerical information before that time?

I propose that the original format was natural frequencies, acquired by natural sampling. Let me explain what this means by a parallel to the mammography problem, using the same numbers. Think about a physician in an illiterate society. Her people have been afflicted by a new, severe disease. She has no books nor statistical surveys; she must rely solely on her experience. Fortunately, she 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, eight showed the symptom; of the 990 not afflicted, 95 did. Thus, there were 8 + 95 = 103 people who showed the symptom, and only 8 of these had the disease. Now a new patient appears. He has the symptom. What is the probability that he actually has the disease?

The physician in the illiterate society does not need a pocket calculator to estimate the Bayesian posterior probability. All she needs to do is to keep track of the number of symptom and disease cases (8) and the number of symptom and no disease cases (95). The probability that the new patient actually has the disease can be “seen” easily from these frequencies:

$$P(H|D) = \frac{a}{a + b} = \frac{8}{8 + 95} \quad (2)$$

Equation 2 is Bayes’s rule for natural frequencies, in which $a$ is the number of cases with symptom and disease, and $b$ is the number of cases having the symptom but lacking the disease. The chance that the new patient has the disease is less than 8 out of 100, or 8%. Our physician who learns from experience cannot be fooled as easily into believing that the chances are about 75%, as many of her contemporary colleagues did.

The comparison between Equations 1 and 2 reveals an important theoretical result: Bayesian reasoning is computationally simpler (in terms of the number of operations performed, such as additions and multiplications) when the information is in natural frequencies (Equation 2) rather than in probabilities (Equation 1) (see Kleiter, 1994). Incidentally, as Equation 2
shows, the base rates of event frequencies (such as 10 in 1,000) need not be kept in memory; they are implicit in the two frequencies, \(a\) and \(b\).

Let me be clear how the terms “natural sampling” and “natural frequencies” relate (Gigerenzer & Hoffrage, 1995). Natural sampling is the sequential process of updating event frequencies from experience. A foraging organism who, day after day, samples potential resources for food and learns the frequencies with which a cue (e.g., the presence of other species) predicts food, performs natural sampling by updating the frequencies \(a\) and \(b\) from observation to observation. Natural sampling is different from systematic experimentation, in which the sample sizes (the base rates) of each treatment group are fixed in advance. For instance, in a clinical experiment, one might select 100 patients with cancer and 100 without cancer, and then perform tests on these groups. By fixing the base rates, the frequencies obtained in such experimental designs no longer carry information about the base rates. This is not to say that controlled sampling in systematic experiments is useless; it just serves a different purpose. Brunswik’s (1955) method of “representative sampling” in a natural environment is an example of applying the idea of natural sampling to experimental design.

Natural frequencies report the final tally of a natural sampling process. There is more than one way to present the final tally. In the case of the physician in the illiterate society, I specified the total number of observations (1,000), the frequency of the disease, and the frequencies \(a\) and \(b\) of hits and false positives, respectively: “In her lifetime, she has seen 1,000 people, 10 of whom had the disease. Of those 10, eight showed the symptom; of the 990 not afflicted, 95 did.” This is a straightforward translation of the base rates, hit rates, and false positive rates into natural frequencies. Alternatively, one can communicate the frequencies \(a\) and \(b\) alone: “In her lifetime, she has seen 8 people with symptom and disease, and 95 people with symptom and no disease.” The former natural frequencies use a standard menu (“standard” because slicing up the information in terms of base rate, hit rate, and false positive rate is deeply entrenched today), the latter use a short menu (Gigerenzer & Hoffrage, 1995). Both lead to the same result.

Natural frequencies must not to be confused with a representation in terms of relative frequencies (e.g., a base rate of .01, a hit rate of .79, and a false positive rate of .096). Relative frequencies are, like probabilities and percentages, normalized numbers that no longer carry information about the natural base rates (Gigerenzer & Hoffrage, 1995). Relative frequencies, probabilities, and percentages are to human reasoning algorithms (that do Bayesian-type inference) like sodium vapor lamps to human color-constancy algorithms. This analogy has, like every analogy, its limits. For instance, humans can be taught, although with some mental agony, to reason by probabilities, but not, I believe, to maintain color constancy under sodium vapor illumination.

Note that the total number of observations—communicated only when natural frequencies are expressed in the standard menu—need not be the actual total number of observations. It can be any convenient number
such as 100 or 1,000. The computational simplicity of natural frequencies holds independently of whether the actual or a convenient number is used. For example, if the actual sample size was 5,167 patients, one can nevertheless represent the information in the same way as above. "For every 1,000 patients we expect 10 who have breast cancer, and 8 out of these 10 will test positive."

The hypothesis that mental algorithms were designed for natural frequencies is consistent with (a) a body of studies that report that humans can monitor frequencies fairly accurately (Barsalou & Ross, 1986; Hintzman & Block, 1972; Jonides & Jones, 1992), (b) the thesis that humans process frequencies (almost) automatically, that is, without or with little effort, awareness, and interference with other processes (Hasher & Zacks, 1984), (c) the thesis that probability learning and transfer derive from frequency learning (Estes, 1976), and (d) developmental studies on counting in children and animals (e.g., Gallistel & Gelman, 1992). This is not to say that humans and animals count all possible events equally well, nor could they. A conceptual mechanism must first decide what the units of observation are so that a frequency encoding mechanism can count them. This preceding conceptual process is not dealt with by the hypothesis that mental algorithms are designed for natural frequencies (but see the connection proposed by Brase, Cosmides, & Tooby, in press).

Thus, my argument has two parts: evolutionary (and developmental) primacy of natural frequencies, and ease of computation. First, mental algorithms, from color constancy to inductive reasoning, have evolved in an environment with fairly stable characteristics. If there are mental algorithms that perform Bayesian-type inferences from data to hypotheses, these are designed for natural frequencies acquired by natural sampling, and not for probabilities or percentages. Second, when numerical information is represented in natural frequencies, Bayesian computations reduce themselves to a minimum. Both parts of the argument are necessary. For instance, the computational part could be countered by hypothesizing that there might be a single neuron in the human mind that almost instantaneously computes Equation 1 on the basis of probability information. The evolutionary part of the argument makes it unlikely that such a neuron has evolved that computes using an information format that was not present in the environment in which our ancestors evolved.

Predictions

This argument has testable consequences. First, laypeople—that is, persons with no professional expertise in diagnostic inference—are more likely to reason the Bayesian way when the information is presented in natural frequencies than in a probability format. This effect should occur without any instruction in Bayesian inference. Second, experts such as physicians who make diagnostic inferences on a daily basis should, despite their experience, show
the same effect. Third, the “inevitable illusions” (Piattelli-Palmarini, 1994), such as base rate neglect should become evitable by using natural frequencies. Finally, natural frequencies should provide a superior vehicle for teaching Bayesian inference. In what follows, I report tests of these predictions and several examples drawn from a broad variety of everyday situations.

This is not to say that probabilities are useless or perverse. In mathematics, they play their role independent of whether or not they suit human reasoning, just as Riemannian and other non-Euclidean geometries play their roles independent of the fact that human spatial reasoning is Euclidean.

Breast Cancer

Eddy (1982) provides only a scant, one-paragraph description of his study of physicians’ intuitions, and refers to a study by Casscells, Schoenberger, and Grayboys (1978) with similar results. Both studies used a probability format. Would natural frequencies make any difference to experts such as physicians? Ulrich Hoffrage and I tested 48 physicians in Munich, Germany on the mammography problem. These physicians had an average professional experience of 14 years. Twenty-four physicians read the information in a probability format as in Eddy’s study, the other 24 read the same information in natural frequencies. Physicians were always asked for a single-event probability (as in Eddy’s study) when the information was in probabilities; they were always asked for a frequency judgment when the information was in natural frequencies. The two formats of the mammography problem are shown in Table 1. Each physician got four diagnostic problems (including the mammography problem), two in a probability format and two in natural frequencies (the details are in Gigerenzer, 1996b; Gigerenzer & Hoffrage, forthcoming; Hoffrage & Gigerenzer, 1996).

In the probability format, only 2 out of 24 physicians (8%) came up with the Bayesian answer. The median estimate of the probability of breast cancer after a positive mammogram was 70%, consistent with Eddy’s findings. With natural frequencies, however, 11 out of 24 physicians (46%) responded with the Bayesian answer. Across all four diagnostic problems, similar results were obtained—10% Bayesian responses in the probability format, and 46% with natural frequencies.

Natural frequencies also changed the physicians’ non-Bayesian inferences. When information was in the form of probabilities, the two dominant non-Bayesian strategies consisted of subtracting the false positive rate from the sensitivity, or simply taking the sensitivity. Both strategies ignore base rates. With natural frequencies, however, these two strategies largely disappeared, and physicians’ dominant non-Bayesian strategies focused exclusively on base rates—the base rate of the disease, or the base rate of a positive test. Natural frequencies not only changed physicians’ reasoning but made them also feel less nervous, more relaxed, and in need of less time to complete the task.
Table 1.1 The Mammography Problem: Probability Format and Natural Frequencies

To facilitate early detection of breast cancer, women are encouraged from a particular age on to participate at regular intervals in routine screening, even if they have no obvious symptoms. Imagine you use mammography to conduct such a breast cancer screening in a certain region. For symptom-free women age 40 to 50 who participate in screening using mammography, the following information is available for this region:

**Probability format**

The probability that one of these women has breast cancer is 1%.

If a woman has breast cancer, the probability is 80% that she will have a positive mammogram.

If a woman does not have breast cancer, the probability is 10% that she will still have a positive mammogram.

Imagine a woman (age 40 to 50, no symptoms) who has a positive mammogram in your breast cancer screening. What is the probability that she actually has breast cancer? _____% 

**Natural frequencies**

Ten out of every 1,000 women have breast cancer.

Of these 10 women with breast cancer, 8 will have a positive mammogram.

Of the remaining 990 women without breast cancer, 99 will still have a positive mammogram.

Imagine a sample of women (age 40 to 50, no symptoms) who have positive mammograms in your breast cancer screening. How many of these women do actually have breast cancer? _____ out of _____.

---

We obtained essentially identical results when we tested lay people (Gigerenzer & Hoffrage, 1995). Lay people and experienced physicians were equally helpless with probabilities, and did not by themselves spontaneously translate probability information into natural frequencies. Some even retranslated frequencies into percentages, because they believed that doing so was the only right way to represent uncertainty. A remarkable result was that when students worked on 30 problems in which natural frequencies and probability formats alternated randomly from problem to problem, students continued to fail on the probability formats and to solve the frequency formats at about the same rate, with little spontaneous transfer (Gigerenzer & Hoffrage, 1995).

Even those who are experienced with statistics can have problems "seeing" through probabilities as easily as through frequencies. Colleagues who work with Bayes's rule on a daily basis often falter when confronted with a specific problem to be solved on the spot. I grant that few people are skilled at mental arithmetic under any circumstances, but it is nevertheless note-
worthy that experts as well as laymen seem to do better when calculations involve natural frequencies rather than probabilities.

The lesson of these results is not to blame physicians' or students' minds when they stumble over probabilities. Rather, the lesson is to represent information in textbooks, in curricula, and in physician-patient interactions in natural frequencies that correspond to the way information was encountered in the environment in which human minds evolved.

**Colon Cancer**

The fecal occult blood test is a widely used and well-known test for colon cancer. Windeler and Köbberling (1986) report that while physicians overestimated the (posterior) probability that a patient has colon cancer if the fecal occult blood test is positive, they also overestimated the base rate of colon cancer, the sensitivity (hit rate), and the false positive rate of the test. Windeler and Köbberling asked these physicians about probabilities and percentages. Would natural frequencies improve physicians' estimates of what a positive test tells about the presence of colon cancer? The 48 physicians in the study reported above were given the best available estimates for the base rate, sensitivity, and false positive rate, as published in Windeler and Köbberling (1986). The following is a shortened version of the full text (structured like the mammography problem in Table 1) given to the physicians. In the probability format, the information was:

The probability that a person has colon cancer is 0.3%.

If a person has colon cancer, the probability that the test is positive is 50%.

If a person does not have colon cancer, the probability that the test is positive is 3%.

What is the probability that a person who tests positive actually has colon cancer?

When one inserts these values into Bayes's rule (Equation 1), the resulting probability is 4.8%. In natural frequencies, the information was:

30 out of every 10,000 people have colon cancer.

Of these 30 people with colon cancer, 15 will test positive.

Of the remaining 9,970 people without colon cancer, 300 will still test positive.

Imagine a group of people who test positive. How many of these will actually have colon cancer?

When the information was in the probability format, only 1 out of 24 physicians (4%) could find the Bayesian answer, or anything close to it. The median estimate was one order of magnitude higher, namely 47%. When the information was presented in natural frequencies, 16 out of 24 physicians (67%) came up with the Bayesian answer (details are in Gigerenzer, 1996b; Hoffrage & Gigerenzer, 1996).
Wife Battering

Alan Dershowitz, the Harvard law professor who advised the defense in the first O. J. Simpson trial, claimed repeatedly that evidence of abuse and battering should not be admissible in a murder trial. In his best-seller, *Reasonable Doubts: The Criminal Justice System and the O. J. Simpson Case* (1996), Dershowitz says: “The reality is that a majority of women who are killed are killed by men with whom they have a relationship, regardless of whether their men previously battered them. Battery, as such, is not a good independent predictor of murder” (p. 105). Dershowitz stated on U.S. television in March 1995 that only about a tenth of 1% of wife batterers actually murder their wives. In response to Dershowitz, I. J. Good, a distinguished professor emeritus of Statistics at the Virginia Polytechnic Institute, published an article in *Nature* to correct for the possible misunderstandings of what that statement implies for the probability that O. J. Simpson actually murdered his wife in 1994 (Good, 1995). Good’s argument is that the relevant probability is not the probability that a husband murders his wife if he batter her. Instead, the relevant probability is the probability that a husband has murdered his wife if he battered her and if she was actually murdered by someone. More precisely, the relevant probability is not $p(G|Bat)$ but $p(G|Bat\ and\ M)$, in which $G$ stands for “the husband is guilty” (that is, did the murder in 1994), $Bat$ means that “the husband battered his wife,” and $M$ means that “the wife was actually murdered by somebody in 1994.”

My point concerns the way Good presents his argument, not the argument itself. Good presented the information in single-event probabilities and odds (rather than in natural frequencies). I will first summarize Good’s argument as he made it. I hope I can demonstrate that you the reader—unless you are a trained statistician or exceptionally smart with probabilities—will have some difficulty following it. Thereafter, I will present the same argument in natural frequencies, and confusion should turn into insight. Let’s see.

Good bases his calculations of $p(G|Bat\ and\ M)$ on the odds version of Bayes’s rule:

$$\frac{p(G|Bat\ and\ M)}{p(G|Bat\ and\ \bar{M})} = \frac{p(G|Bat) \times p(M|G \ and \ Bat)}{p(G|Bat) \times p(M|G \ and \ Bar)}$$

which, in the present case is:

$$\frac{p(G|Bat\ and\ M)}{p(G|Bat\ and\ \bar{M})} = \frac{p(G|Bat) \times p(M|G \ and \ Bat)}{p(G|Bat) \times p(M|G \ and \ Bar)} \quad (3)$$

where $\bar{G}$ stands for “the husband is not guilty.”

The following six equations (Good-1 to Good-6) show Good’s method of explaining to the reader how to estimate $p(G|Bat\ and\ M)$. Good starts with Dershowitz’s figure of a tenth of 1%, arguing that if the husband commits the murder, the probability is at least 1/10 that he will do it in 1994:

$$p(G|Bat) > (1/10) \ (1/1,000) = 1/10,000 \quad (Good-1)$$

Therefore, the prior odds (O) are:

$$O(G|Bat) > 1/9,999 = 1/10,000 \quad (Good-2)$$
Furthermore, the probability of a woman being murdered given that her husband has murdered her (whether he is a batterer or not) is unity:

$$p(M|G\text{ and } \text{Bat}) = p(M|G) = 1$$  \hspace{1cm} (Good-3)

Because there are about 25,000 murders per year in the US population of about 250,000,000, Good estimates the probability of a woman being murdered, but not by her husband, as:

$$p(M|\bar{G}\text{ and } \text{Bat}) = p(M|\bar{G}) = 1/10,000$$  \hspace{1cm} (Good-4)

From Equations Good-3 and Good-4, it follows that the likelihood ratio is about 10,000/1; therefore, the posterior odds can be calculated:

$$O(G|M\text{ and } \text{Bat}) > 10,000/10,000 = 1$$  \hspace{1cm} (Good-5)

That is, the probability that a murdered, battered wife was killed by her husband is:

$$p(G|\text{Bat and } M) > 1/2$$  \hspace{1cm} (Good-6)

Good’s point is that “most members of a jury or of the public, not being familiar with elementary probability, would readily confuse this with \(P(G|\text{Bat})\), and would thus be badly misled by Dershowitz’s comment” (Good, 1995, p. 541). He adds that he sent a copy of this note to both Dershowitz and the Los Angeles Police Department, reminding us that Bayesian reasoning should be taught at the pre-college level.

Good’s persuasive argument, I believe, could have been understood more easily by his readers and the Los Angeles Police Department if the information had been presented in natural frequencies rather than in the single-event probabilities and odds in the six equations. As with breast cancer and colon cancer, one way to represent information in natural frequencies is to start with a concrete sample of individuals divided into subclasses, in the same way it would be experienced by natural sampling. Here is a frequency version of Good’s argument.

**Good’s Argument in Natural Frequencies**

Think of 10,000 battered married women. Within one year, at least one will be murdered by her husband. Of the remaining 9,999 who are not killed by their husbands, one will be murdered by someone else. Thus we expect at least two battered women to be murdered, at least one by her husband, and one by someone else. Therefore, the probability \(p(G|\text{Bat and } M)\) that a murdered, battered woman was killed by her husband is at least ½.

This probability is not to be confounded with the probability that O. J. Simpson is guilty; a jury must take into account much more evidence than battering. But the probability shows that abuse-and-battering is a good predictor of the husband’s (or boyfriend’s) guilt, disproving Dershowitz’s assertion to the contrary.

In natural frequencies, Good’s argument is short and transparent. My conjecture is that more ordinary people, including employees of the Los
Angeles Police Department and jurors, could understand and commu-
nicate the argument if the information were represented in natural frequen-
cies rather than in probabilities or odds.

In legal jargon, evidence of wife battering is probative, not prejudicial. This analysis is consistent with the impressive transcultural evidence about homicide accumulated by Daly and Wilson (1988). The typical function of wife battering seems to be to exert proprietary rights over the sexuality and reproduction of women, as well as threats against infidelity. Battering can "spill over" into killing, and killing is the tip of a huge iceberg of wife abuse.

**AIDS Counseling**

Under the headline, "A False HIV Test Caused 18 Months of Hell," the Chicago Tribune (3/5/93) published the following letter and response:

> Dear Ann Landers: In March 1991, I went to an anonymous testing center for a routine HIV test. In two weeks, the results came back positive.
> I was devastated. I was 20 years old and doomed. I became severely depressed and contemplated a variety of ways to commit suicide. After encouragement from family and friends, I decided to fight back.
> My doctors in Dallas told me that California had the best care for HIV patients, so I packed everything and headed west. It took three months to find a doctor I trusted. Before this physician would treat me, he insisted on running more tests. Imagine my shock when the new results came back negative. The doctor tested me again, and the results were clearly negative.
> I'm grateful to be healthy, but the 18 months I thought I had the virus changed my life forever. I'm begging doctors to be more careful. I also want to tell your readers to be sure and get a second opinion. I will continue to be tested for HIV every six months, but I am no longer terrified.
> David in Dallas

> Dear Dallas: Yours is truly a nightmare with a happy ending, but don't blame the doctor. It's the lab that needs to shape up. The moral of your story is this: Get a second opinion. And a third. Never trust a single test. Ever.
> Ann Landers

David does not mention what his Dallas doctors told him about the chances that he actually had the virus after the positive test, but he seems to have inferred that a positive test meant that he had the virus, period. In fact, when we studied AIDS counselors in Germany, we found that many doctors and social workers (erroneously) tell their low-risk clients that a positive HIV test implies that the virus is present (Gigerenzer, Hoffrage, & Ebert, in press). These counselors know that a single ELISA (enzyme-linked immunosorbent assay) test can produce a false positive, but they erroneously assume that the whole series of ELISA and Western blot tests would wipe out every false positive. How could a doctor have explained the actual risk to David and spared him the nightmare?
I do not have HIV statistics for Dallas, so I will use German figures for illustration. (The specific numbers are not the point here.) In Germany, the prevalence of HIV infections in heterosexual men between the ages of 20 and 30 who belong to no known risk group can be estimated as about 1 in 10,000, or 0.01%. The corresponding base rate for homosexual men is estimated at about 1.5%. The hit rate (sensitivity) of the typical test series (repeated ELISA and Western blot tests) is estimated at about 99.8%. The estimates of the false positive rate vary somewhat; a reasonable estimate seems to be 0.01% (Gigerenzer, Hoffrage, & Ebert, in press). Given these values, and assuming that David was at the time of the routine HIV test a heterosexual man with low-risk behavior, what is the probability that he actually had the virus after testing positive? If his physician had actually given David these probabilities, David nevertheless might not have understood what to conclude.

But the physician could have communicated the information in natural frequencies. He might have said, “Your situation is the following: Think of 10,000 heterosexual men like you. We expect one to be infected with the virus, and he will, with practical certainty, test positive. From the 9,999 men who are not infected, one additional individual will test positive. Thus we get two individuals who test positive, but only one of them actually has the virus. This is your situation. The chances that you actually have the virus after the positive test are about 1 in 2, or 50%.” If the physician had explained the risk in this way, David might have understood that there was, as yet, no reason to contemplate suicide or to move to California.

If David were a member of a risk group, say a homosexual with a 1.5% base rate of HIV infection, the estimate would have been different. Here, the physician might have explained, “Think of 10,000 homosexual men. We expect 150 to be infected with the virus, and they all will likely test positive. From the 9,850 men who are not infected, we expect that one other will test positive. Thus we have 151 men who test positive, and 150 of these have the virus. Your chances of not having the virus are 1 out of 151, that is, less than 1%.” This would not be good news, but it would still be better than leaving the doctor’s office with the belief that an HIV infection is absolutely certain. David might be the lucky one, and his odds are certainly better than winning a lottery.

We do not know what risk group David was in. Whatever the statistics are, however, most people of average intelligence can understand the risk of HIV after a positive test when the numbers are represented by a counselor in natural frequencies. Not one of the 20 AIDS counselors studied by Gigerenzer, Hoffrage, & Ebert (in press), however, explained his or her client’s risk in natural frequencies. Except for the prevalence of HIV, all numerical information was communicated to the client in percentages.

Ann Landers’s answer—don’t blame the doctor, blame the lab—however, overlooks the fact that despite whatever possible reasons there may be for false positives (such as the presence of cross-reacting antibodies or blood
samples being confused in the lab), a doctor should inform the patient that false positives occur, and about how frequently they occur.

**Expert Witnesses**

Evidentiary problems such as the evaluation of eyewitness testimony constituted one of the first domains of probability theory (Gigerenzer et al., 1989, chap. 1). Statisticians have taken the stand as expert witnesses for almost a century now: In the Dreyfus case in the late nineteenth century in France, or more recently, in *People vs. Collins* in California (Gigerenzer et al., 1989, chap. 7; Koehler, 1992). The convictions in both cases were ultimately reversed and the statistical arguments discredited. Part of the problem seems to have been that the statistical arguments were couched not in frequencies but in probabilities which confused both the prosecution who were making the arguments and the jury and the judges who tried to understand the arguments. I will explain this point with the case of a chimney sweep who was accused of having committed a murder in Wuppertal, Germany (Schrage, n.d.).

The *Rheinische Merkur* (No. 39, 1974) reported:

On the evening of July 20, 1972, the 40-year-old Wuppertal painter Wilhelm Fink and his 37-year-old wife Ingeborg took a walk in the woods and were attacked by a stranger. The husband was hit by three bullets in the throat and the chest, and fell down. Then the stranger attempted to rape his wife. When she defended herself and unexpectedly, the shot-down husband got back on his feet to help her, the stranger shot two bullets into the wife’s head and fled.

Three days later, a forest ranger discovered 20 kilometers from the scene of the crime the car of Werner Wiegand, a 25-year-old chimney sweep who used to spend his weekends in the vicinity. The husband, who had survived, at first thought he recognized the chimney sweep in a photo. Later, he grew less certain and began to think that another suspect was the murderer. When the other suspect was found innocent, however, the prosecution came back to the chimney sweep and put him on trial. The chimney sweep had no previous convictions and denied being the murderer. The *Rheinische Merkur* described the trial:

After the experts had testified and explained their “probability theories,” the case seemed to be clear: Wiegand, despite his denial, must have been the murderer. Dr. Christian Rittner, a lecturer at the University of Bonn, evaluated the traces of blood as follows: 17.29% of German citizens share Wiegand’s blood group, traces of which have been found underneath the fingernails of the murdered woman; 15.69% of Germans share [her] blood group that was also found on Wiegand’s boots; based on a so-called “cross-combination” the expert subsequently calculated an overall probability of 97.3% that Wiegand “can be considered the murderer.” And concerning the textile fiber traces which were found both on Wiegand’s clothes and on those of the vic-
tim [...] Dr. Ernst Röhm from the Munich branch of the State Crime Department explained: "The probability that textile microfibers of this kind are transmitted from a human to another human who was not in contact with the victim is at most 0.06%. From this results a 99.94% certainty for Wiegand being the murderer."

Both expert witnesses agreed that, with a high probability, the chimney sweep was the murderer. These expert calculations, however, collapsed when the court discovered that the defendant was in his hometown, 100 kilometers away from the scene of the crime at the time of the crime.

So what was wrong with the expert calculations? One can dispel the confusion in court by representing the uncertainties in natural frequencies. Let us assume that the blood underneath the fingernails of the victim was indeed the blood of the murderer, that the murderer carried traces of the victim’s blood (as the expert witnesses assumed), and that there were 10,000,000 men in Germany who could have committed the crime (Schrage, n.d.). Let us assume further that on one of every 100 of these men a close examination would find microscopic traces of foreign blood, that is, on 100,000 men. Of these, some 15,690 men (15.69%) will carry traces from blood that is of the victim’s blood type. Of these 15,690 men, some 2,710 (17.29%) will also have the blood type that was found underneath the victim’s fingernails (here, I assume independence between the two evidences). Thus, there are some 2,710 men (including the murderer) who might appear guilty based on the two pieces of blood evidence. The chimney sweep is one of these men. Therefore, given the two blood evidences the probability that the chimney sweep is the murderer is about 1 in 2,710, and not 97.3%, as the first expert witness testified.

The same frequency method can be applied to the textile traces. Let us assume that the second expert witness was correct when he said that the probability of the chimney sweep carrying the textile trace, if he were not the murderer, would be at most 0.06%. Let us assume as well that the murderer actually carries that trace. Then some 6,000 of the 10,000,000 would carry this textile trace, and only one of them would be the murderer. Thus, the probability that the chimney sweep was the murderer, given the textile fiber evidence, was about 1 in 6,000, and not 99.94%, as the second expert witness testified.

What if one combines both the blood and the textile evidences together, which seems not to have happened at the trial? In this case, one of the 2,710 men who satisfy both blood type evidences would be the murderer, and he would show the textile traces. Of the remaining innocent men, we expect one or two (0.06%) to also show the textile traces (assuming mutual independence of the three evidences). Thus, there would be two or three men who satisfy all three types of evidence. One of them is the murderer. Therefore, the probability that the chimney sweep was the murderer, given the two blood sample evidences and the textile evidence, would be between ⅖ and ⅙. This probability would not be beyond reasonable doubt.
Cognitive Illusions

Frequencies not only make everyday inferences easier, they also tend to make “cognitive illusions” of the laboratory type largely disappear. I have summarized this evidence elsewhere (Gigerenzer, 1991, 1994). One example is a cognitive illusion called “overconfidence bias.” Students were given questions such as, “Which city has more inhabitants: Islamabad or Hyderabad?” They were then asked to estimate the probability (confidence) that their answer was correct. The typical result was that when students said they were 100% confident, they were correct in only about 85% of the cases. When they said they were 90% confident, they were correct in only 75% of the cases, and so on (Lichtenstein, Fischhoff, & Phillips, 1982). This discrepancy between subjective probability and objective frequency was labeled the “overconfidence bias,” and human disasters of many kinds, from deadly accidents in industry to errors in the legal process, have been attributed to that “cognitive illusion.” However, when we replaced the probability judgments by frequency judgments, the apparently stable cognitive illusion disappeared (Gigerenzer, Hoffrage, & Kleinbölting, 1991). We asked students after every 50 questions to estimate how many they got correct, and these frequency judgments no longer overestimated the actual frequencies of correct answers. In fact, they turned out to be fairly accurate, even with a tendency towards underestimation.

A second example is a medical disease problem with which Casscells, Schoenberger and Grayboys (1978) had demonstrated base rate neglect by staff and students at Harvard Medical School. Cosmides and Tooby (1996) replaced the probabilities with natural frequencies, and Bayesian responses increased from 12% (in the original probability format) to 76%. An instruction to visualize frequencies in a grid boosted the performance up to 92%. More generally, natural frequencies reduce the “base rate fallacy” in the cab problem and similar “toy” problems (Gigerenzer, 1994; Gigerenzer & Hoffrage, 1995).

A third and final example is the “Linda” problem. People read a description of Linda that suggests that she is a feminist. Thereafter, the people are asked which is more probable: (a) Linda is a bank teller; or, (b) Linda is a bank teller and active in the feminist movement. Some 80%-90% of subjects usually chose (b), a response which Tversky and Kahneman (1983) labeled the “conjunction fallacy,” because the probability of a conjunction of two events (teller and feminist) cannot be larger than the probability of one of these events (teller). This “conjunction fallacy” in the Linda problem and related tasks, however, largely disappeared when people were asked for judgments of frequencies: Think of 200 women like Linda. How many of them are (a) bank tellers, (b) bank tellers and active in the feminist movement? Replacing probabilities by frequencies made conjunction violations drop from 80–90% to 0–20% (Hertwig & Gigerenzer, 1997; similar results were reported by Fiedler, 1988; Tversky & Kahneman, 1983). The effect of frequency representations and judgments on “cognitive illu-
sions” is the strongest and most consistent “debiasing method” known today.

Teaching Statistical Reasoning

The teaching of statistical reasoning is, like that of reading and writing, part of forming an educated citizenship. Our technological world, with its abundance of statistical information, makes the art of dealing with uncertain information particularly relevant. Reading and writing is taught to every child in modern Western democracies, but statistical thinking is not (Shaughnessy, 1992). The result has been termed “innumeracy” (Paulos, 1988). But can statistical reasoning be taught? Previous studies that attempted to teach Bayesian inference, mostly by corrective feedback, had little or no training effect (e.g., Peterson, DuCharme, & Edwards, 1968; Schaefer, 1976). This result seems to be consistent with the view that the mind does not naturally reason the Bayesian way. However, the argument developed in this chapter suggests a “natural” method of teaching: instruct people how to represent probability information in natural frequencies. Recall that students and physicians alike did not do this spontaneously, with very few exceptions (e.g., Gigerenzer & Hoffrage, 1995).

Peter Sedlmeier and I designed a tutorial program that teaches Bayesian reasoning, based on the assumption that cognitive algorithms have evolved for dealing with natural frequencies (Sedlmeier, 1997; Sedlmeier & Gigerenzer, 1998). The goal of this tutorial is to teach participants how to reason the Bayesian way when the information is represented in probabilities, as is usually the case in newspapers, medical textbooks, and other information sources. The computerized tutorial instructs participants in how to represent the probability information in terms of natural frequencies, rather than teaching them how to insert probabilities into Bayes’s rule (Equation 1). The tutorial consists of two parts. In the first part, participants are shown how to translate probability information into natural frequencies, visually aided by a frequency tree (or a frequency grid); the method is illustrated by two medical problems, one of them the mammography problem. In the second part, participants solve eight other problems, with step-by-step guidance on what to do as well as step-by-step feedback. If participants have difficulties, the system provides immediate help that ensures that every participant solves all problems correctly. We conducted an evaluation study with four groups: two groups were taught how to represent probabilities in natural frequencies (visually demonstrated by a frequency tree and a frequency grid, respectively), one control group was taught how to insert probabilities into Bayes’s rule using a similar computer tutorial (“rule training”), and a second control group received no training.

Sixty-two University of Chicago students participated in the study. In the rule-training group, the median number of Bayesian solutions increased from 0% (pretraining baseline) to 35% after training, and the values for the
no-training control increased slightly from 0% to 5%. In the two groups that learned how to construct frequency representations by constructing frequency trees or frequency grids, the median number of Bayesian solutions increased from 0% and 5% to 80% and 70%, respectively. Thus, the immediate success of the frequency tutorials was about twice as high as that of the rule training. But did the performance last over time, or was it subject to the usual steep forgetting curve following a successful test? In a five-week follow up, the median performance of the rule-training group was down to a mere 15%. The median performance of each of the two frequency representation groups five weeks after training, however, was a strong 90%.

Thus, there is evidence that (what I take to be) the natural format of information in the environment in which humans evolved can be used to teach people how to deal with probability information. This may be good news both for instructors who plan to design pre-college curricula that teach young people how to infer risks in a technological world, and for those unfortunate souls among us charged with teaching undergraduate statistics.

Conclusions

Information needs representation. If a representation is recurrent and stable during human evolution, one can expect that mental algorithms are designed to operate on this representation. In this chapter, I applied this argument to the understanding of human inferences under uncertainty. The thesis is that mental algorithms were designed for natural frequencies, the recurrent format of information until very recently. I have dealt with a specific class of inferences that correspond to a simple form of Bayesian inferences, in which one of several possible states is inferred from one or a few cues. Here, mental computations are simpler when information is encountered in the same form as in the environment in which our ancestors evolved, rather than in the modern form of probabilities or percentages. The evidence from a broad variety of everyday situations and laboratory experiments shows that natural frequencies can make human minds smarter.

Notes

I would like to thank Ralph Hertwig, Ulrich Hoffrage, James S. Magnuson, Laura Martignon, and Anita Todd for their helpful comments.

1. As we know from Stephen M. Stigler's Law of Eponymy, no scientific discovery is named after its original discoverer, and Bayes's rule seems to be no exception to this rule (Stigler, 1983).

2. For a critical discussion of these interpretations see Cohen (1981), Gigerenzer (1994, 1996a), Gigerenzer and Murray (1987, chap. 5), and Lopes (1991); for a reply, see Kahneman and Tversky (1996).

3. However, there is a price to be paid if one replaces the actual with a convenient sample size. One can no longer compute second-order probabilities (Kleiter, 1994).
4. Good possibly assumed that the average wife batterer is married less than 10 years. Good also made a second calculation assuming a value of $p(G|Bat)$ that is half as large.

References


Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–162). New York: Wiley.

