only a belief about a category's probabilistic attributes, but an *incorrect* one. By definition, then, the effect of this belief was harmful, and any influence it at all had was too much.

Base rates researchers started with normative models of decision making based on Bayesian theory and, in some cases, urn models. As Koehler points out, a real world problem often glossed over in research is how to define and then determine the base rate that is relevant to a given decision. Typical experiments simply presented this information in the protocol, as in Kahneman and Tversky's (1973) famous bald statement that the target person comes from a population of 30 engineers and 70 lawyers. The information might as well come from God. It is treated in this research (by the researchers if not by the subjects) as definitionally true, so its effect on judgment can only be beneficial, and any failure to use it full-strength can only harm accuracy.

This leads to a final question – Is it really good, as base rate literature would have it, or bad, as stereotypes literature would have it, to use one's preexisting beliefs about category attributes when forming judgments about a member of a category? The answer is neither. One should use this information when it is accurate and not use it when it is inaccurate. Research to ascertain the accuracy of stereotypes, or base rates, is outside the framework of normative decision-making research as described by Koehler.

What is needed, in addition to the suggestions for more "ecologically valid" research (sect. 5; the Brunswikian term is "representative"), is research on the beliefs people hold about social (and other) categories, as well as research that examines, through converging evidence, the degree to which these beliefs are correct or mistaken. For an example see the article by Swim (1994), comparing popular beliefs about the attributes of men and women with meta-analytic conclusions about actual gender differences.

Content-oriented research complements the process-oriented research described by Koehler in an important way. It examines not just how information is used, but is realistic accuracy (Funder, in press, a). The effect of base rate, stereotypic, or any other kind of information on judgmental accuracy, depends only in part on how that information is used. It also depends on whether the information was any good in the first place.

Why do frequency formats improve Bayesian reasoning? Cognitive algorithms work on information, which needs representation

Gerd Gigerenzer

Max Planck Institute for Psychological Research, 80802 Munich, Germany.
giger@mpipf-muenchen.mpg.de

Abstract: In contrast to traditional research on base-rate neglect, an ecologically-oriented research program would analyze the correspondence between cognitive algorithms and the nature of information in the environment. Bayesian computations turn out to be simpler when information is represented in frequency formats as opposed to the probability formats used in previous research. Frequency formats often enable even uninstructed subjects to perform Bayesian reasoning.

Some years ago I applied for an American green card. The U.S. immigration office demanded an HIV test, and I was told that a positive test (actually, a positive test confirmed by a second one) would result in denial of the green card. The morning I drove to the U.S. consulate in Frankfurt to take the test, I asked myself what was the probability of having the virus if the test came out positive. At that time I had the following information. About .02% of German men have the HIV virus (base rate). If someone has the virus, there is a 99% chance that the test will be positive (hit rate). If someone does not have the virus, there is still a .5% chance that the test will be positive (false alarm rate). Question – What is the probability of having the virus if the test comes out positive?

One way to answer this question is to take a pencil and paper

and to compute the posterior probability with Bayes' theorem – but I was driving. A less cumbersome method is to change the representation of the information from probabilities (percentages) into absolute frequencies (natural numbers), which can be done while driving. Imagine a sample of 10,000 men. Two have the virus (base rate), and these two will most likely test positive (hit rate). Out of the 9,998 who do not have the virus, some 50 will also test positive (false alarm rate). So we have 52 who test positive. Question – How many of those who test positive actually have the virus?

With this frequency representation, one does not need a pencil and paper or a calculator. The answer can immediately be "seen." About 2 out of the 52 men who test positive have the virus. This figure corresponds to a .04 posterior probability. Note that the information representations, probability and frequency, are mathematically equivalent: they can be mapped onto each other in a one-to-one fashion. But what is equivalent for mathematics may not be equivalent for the mind. Most people have no idea what to do with such information when it is represented as probabilities or percentages: many, however, show insight when the information comes in natural numbers (for details see Gigerenzer & Hoffrage 1995).

Why do frequency formats improve Bayesian reasoning without instruction? Bayesian computations are simpler with frequency formats than with probabilities or percentages. Let the symbols H and $-H$ stand for the two hypotheses (virus or no virus), and D for the potential data (positive HIV test). A Bayesian algorithm for computing the posterior probability $p(H|D)$ with the values given in the probability version amounts to solving the following equation:

$$p(H \mid D) = \frac{p(H)p(D \mid H)}{p(H)p(D \mid H) + p(-H)p(D \mid -H)} \quad (1)$$

In contrast, a Bayesian algorithm for computing the posterior probability $p(H \mid D)$ in the frequency version, requires solving the following equation:

$$p(H \mid D) = \frac{a}{a + b} \quad (2)$$

where a is the number of cases who have both the symptom and the disease, and b is the number of cases having the symptom but lacking the disease. Thus, when information is presented in natural numbers, the Bayesian computations are much simpler than with probabilities or percentages.

The general point is that algorithms need information, and information needs to be represented. Thus, cognitive algorithms ought to be studied in tandem with the external representations of information on which they operate (Marr 1982). This psychological point has been overlooked in most of the research on the so-called base rate fallacy, where the external representation of information has been a matter of convention, not theory. The link between algorithm and external representation is just as important for an electronic calculator. My pocket calculator has an algorithm for multiplication, which is designed for Arabic numbers as input. If I enter binary numbers, garbage comes out. But I cannot conclude from the garbage that my calculator has no algorithm for multiplication. Similarly, one cannot conclude, as has been done in prior research on base rate neglect, that people who come up with wild posterior probabilities do not have a Bayesian algorithm in their mind.

What happens to people's reasoning when researchers use frequency rather than probability representations for problems used in previous research, such as the cab problem and the mammography problem? In one of the largest studies ever done on Bayesian inference (Gigerenzer & Hoffrage 1995; see also Cosmides & Tooby, in press), we found that with frequency representations, subjects arrived at the numerically exact estimate using a Bayesian algorithm (including pictorial equivalents and shortcuts) in about 50% of the cases.

Koehler has rightly pointed out that research on statistical thinking needs to move toward what he calls an ecologically valid research program, and the above analysis of the role of external representation in reasoning elaborates on this vision. A first step toward ecologically minded research is to think about how information has been represented and encountered by humans during most of our history, and to look for cognitive algorithms that are tuned to those representations. From such ecological considerations we might expect cognitive algorithms to be designed for absolute frequencies rather than for probabilities and percentages, which depend on the development of literacy and numeracy (Gigerenzer et al. 1989). An important consequence is that base rates need not be attended to in natural sampling of frequencies (Kleiter 1994). This can be seen from equation 2, in which the base rates are already embodied in the two absolute frequencies. The only information that needs to be monitored are these two frequencies, for example, the number of cases with symptom and with disease and the number of cases with symptom and no disease.

Representation of information in terms of frequencies improves Bayesian inferences without instruction. More generally, frequency representations make various "cognitive illusions" in statistical reasoning largely disappear (for an overview see Gigerenzer 1991; 1994). These results are of course good news for those who would like to believe in some sort of human rationality, for those biologically minded people who wonder how a species so bad at judgment under uncertainty could have survived so long, and for those unfortunate souls charged with teaching undergraduate statistics.

Koehler's target article on the base rate fallacy is timely. Research on the use of base rates has been driven by the formal theory of probability, which is mute about representation and content (but see how Birnbaum, 1983, tailored statistical models to the content of the cab problem). Research has also ignored the problems with base rates that arise when Bayes' theorem is applied to everyday life (e.g., Daston 1988; Earman 1992). For example, while driving to the US consulate for my HIV test, I wondered from which reference class to take the base rate, since I am a member of several reference classes with different base rates. Selecting a reference class introduces a source of subjectivity that statisticians such as Fisher (1935) have held against the routine application of Bayes' theorem.

Ironically, most researchers who hold subjects to Bayesian standards do not themselves adhere to those standards in their research. Experimenters habitually use Fisher's significance testing, not Bayes' theorem, to infer whether subjects reason the Bayesian way. Thus they accuse subjects of committing the base rate fallacy at the same time that they neglect base rates. This is a double standard that should make us suspicious of the mechanical application of norms, Bayesian or otherwise, to evaluating human inference.

Judgment under uncertainty: Evolution may not favor a probabilistic calculus

Lev R. Ginzburg,[a] Charles Janson,[a] and Scott Ferson[b]

[a]*Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794.*
[b]*Applied Biomathematics, Setauket, NY 11733.* **risk@life.bio.sunysb.edu**

Abstract: The environment in which humans evolved is strongly and positively autocorrelated in space and time. Probabilistic judgments based on the assumption of independence may not yield evolutionarily adaptive behavior. A number of "faults" of human reasoning are not faulty under fuzzy arithmetic, a nonprobabilistic calculus of reasoning under uncertainty that may be closer to that underlying human decision making.

Koehler questions whether Bayesian probability theory is the sole correct normative standard against which human statistical judg-

ments should be compared. We think there are evolutionary explanations of why humans might not generally reason by the rules of probability theory. We review some of these and suggest there is a distinct nonprobabilistic calculus that humans may have evolved for making decisions under uncertainty.

Probability theory is almost paradoxically nonintuitive. Aside from the heuristics and biases literature, there are also some famous cases in which even experts have been fooled by seemingly straightforward problems (e.g., the Monty Hall dilemma). Why was probability theory not developed until the 1700s, centuries or even millennia after the flowering of other mathematical subjects that are arguably less germane to everyday decision making, such as geometry and algebra?

In fact, probability theory arose in reaction to games of chance involving artificial gambling devices, such as tossed coins and dice. These devices exhibit dynamical chaos with extreme sensitivity to initial conditions on the time scale of human perception. The devices mimic well the idea of statistical independence, and that explains their usefulness in gambling. In contrast, it is hard to think of examples of independence in natural phenomena, where autocorrelation and other dependencies are the rule. Strong positive correlations are observed in most natural fluctuations of magnitude and time scale relevant to human evolution (Pimm & Redfearn 1988). For example, the best predictor of tomorrow's weather is today's. Even negatively correlated phenomena such as daily, seasonal, and yearly cycles generally reveal positive short-term autocorrelations once the deterministic trend is extracted and variability is expressed as deviations from the mean.

Spatial patterns in ecology also generally exhibit large positive autocorrelation. Early humans would surely learn that a good place to forage for food is near where food was found before. Humans evolving in such autocorrelated landscapes would mainly experience conditions characterized by strong interdependence among events. It is no surprise, then, that human evolution has created decision-making and behavioral patterns that are alienated from probability theory. Similarly, in autocorrelated environments, the results of animal foraging theory can be counterintuitive (Stephens & Krebs 1986, pp. 81–90). Practical application of probability theory virtually demands independence assumptions at some level. Since independence is unusual in the natural world, there may have been little selective pressure for humans to become probabilists.

Arriving at good probabilistic decisions typically requires making use of a body of empirical information. Yet humans seem to be ill-equipped to process frequency information. For example, we are notoriously difficult to educate about the relative risks of environmental carcinogens (Graham 1989), despite the clear importance of the information in daily life. Why should this be? Given the range and depth of other cognitive skills in humans, this inability might be considered surprising.

One possible reason that humans are not much better at collecting, remembering, and using frequency information is that these tasks require time, but indecision is often punished. Taking too long to chose which food to eat may result in getting nothing for dinner. Likewise, upon seeing a predator, one must decide quickly whether to freeze, flee, or fight. The fitness cost of indecision may often exceed the cost of a suboptimal, but quick, decision.

Another possible reason that humans are not better at collecting frequency information is that forgetfulness can be useful. For example, optimality models of foraging suggest that in rapidly changing or patchy environments, food encounter rates should be calculated only over the recent past, ignoring longer-term information (Iwasa et al. 1981). In all realms of behavior, in fact, invariant decisions are a maladaptive strategy whenever circumstances change (e.g., Maynard Smith 1982). In this sense, forgetting the past is an essential component of adjusting behavior to current conditions, which should be strongly favored by evolution.

Despite arguments that probability theory is the only reason-