The major difficulty seems to be that the normative component of a Bayesian model applies only to the way in which the prior probability of a hypothesis should be adjusted according to relevant evidence. Bayes' theorem itself is uninformative in telling us how the prior probability of a hypothesis should be adjudged, or how much credence the evidence should be given prior to any revision of the hypothesis. Those who hold to a fundamentalist creationist theory may be perfect Bayesians, for example, by arguing, as they do, that the fossil record can be interpreted in ways that are consistent with a creationist account. But this does not mean that creationists are not irrational or unscientific in arguing such a case.

Hypotheses, in lay cognition as in science, are typically connected to wider theories, and the strength of belief in the hypothesis is, or should be, associated with the strengths and weaknesses of the wider theory. The wider theory, in turn, in science at least, is appropriately assessed according to several criteria that go beyond the ability of the theory to make correct empirical predictions. For example, scientific theories can be evaluated in terms of their simplicity, their ability to unify disparate phenomena, their logical consistency, and their fertility (Fletcher 1996). None of these criteria can sensibly be derived from a Bayesian perspective, yet all of them may be relevant to an assessment of the rationality or plausibility of the prior probability assigned to a particular hypothesis or theory.

Another problem concerns the normative status of Bayes' theorem itself in evaluating the way that beliefs or hypotheses are revised in the face of new evidence. The pattern of evidence from the base rate literature, and the error and bias research generally, is that people are overconservative (in terms of Bayes' theorem) in revising their prior theories or hypotheses when faced with disconfirmatory evidence. But are people wrong to err on the conservative side? Given the ubiquity of conflicting evidence, and the desirability of retaining a stable view of the world, such theoretical conservatism can plausibly be characterized as normatively appropriate for laypeople and scientists alike.

For stripped-down probabilistic problems that comprise hypotheses and evidence that are (relatively) unencumbered with wider theory, Bayes' theorem may be a fine normative model. However, to treat it as a generic normative model of rational inference, either for lay or scientific domains, is to stretch it well beyond its conceptual resources.

Evolving null hypotheses and the base rate fallacy: A functional interpretation of scientific myth

Brian J. Gibbs

Graduate School of Business, Stanford University, Stanford, CA 94305-5015. **gibbs-brian@gsb.slanford.edu**

Abstract: The meaning of an experimental result depends on the experiment's conceptual backdrop, particularly its null hypothesis. This observation provides the basis for a functional interpretation of belief in the base rate fallacy. On this interpretation, if the base rate fallacy is to be labelled a "myth," then it should be recognized that this label is not necessarily a disparaging one.

Koehler's (f 996t) message regarding the base rate fallacy is right in substance but wrong in tone, and it would be a mistake to take it as reason to lose faith in the field of decision research. It is difficult to argue with much of his incisive analysis. His major themes - that base rate neglect is less pervasive than sometimes asserted, that research should focus on the factors determining when and how much base rates are used, that prescriptive analyses should question the Bayesian normative standard, and that experiments should attempt to capture more of the richness of the natural ecology - all have merit. However, in the course of arguing that the base rate fallacy is a "myth," Koehler's target article may leave the impression that the research focusing on this fallacy should be considered a shameful case of science gone wrong. Quite the contrary, the focus on fallacy in this and other decision contexts can be viewed as following from a reasonable judgment by scientists about what constitutes the most interesting and useful null hypothesis against which to contrast experimental results.

Because data alone are not meaningful, interpretation is an inherent and unavoidable part of science (Kuhn 1970; Peter & Olson 1983). Some of this interpretive activity is explicit, but much of it is implicit, such as when a researcher selects a particular conceptual backdrop against which to view empirical facts. Just as the visual perception of a figure is influenced by background stimuli, the meaning of empirical data is partly determined by the surrounding concepts and ideas relative to which those data are examined. A critical element of an investigation's conceptual backdrop is its null hypothesis, and, as Keren and Thijs (1996, p. 26) rightly point out, due to discretion in selecting null hypotheses, the controversy over the base rate fallacy "boils down to the question of whether the glass is half-full or half-empty." But it should be recognized that the difference between "half-full" and "half-empty" can be important.

In the case of research on base rate use, the conceptual backdrop has often been a normative one, and a favorite null hypothesis has therefore been that subjects will behave like Bayesians. Consequently, mixed results, indicating partial usage of base rates, are interesting not because they show information use, but because they show information neglect. Indeed, given a traditional normative conceptual backdrop with its Bayesian null hypothesis, the information-neglect interpretation *deserves* greater emphasis because it carries greater meaning. Of course, the popularity of the notion of judgment heuristics has shifted the scientific conversation, and when mixed results are viewed relative to a base-rate-fallacy null hypothesis, as they are in parts of Koehler's article, then an information-sensitivity interpretation becomes the important one. This accounts for why in some respects Koehler himself may seem "guilty of the same kind of overstatement of which he accuses past purveyors of the base rate fallacy" (Thomsen & Borgida 1996, p. 39).

This account of belief in the base rate fallacy is not unrelated to the two accounts put forward by Koehler. It shares with his Kuhnian account an emphasis on the importance of the conceptual backdrop, and it shares with his heuristic account a recognition that scientists, too, are limited information processors (if they were not, then they would presumably entertain all possible meanings of a data set by viewing it simultaneously against all possible conceptual backdrops[1]). However, the present account, which explicitly acknowledges the need to construct meaning from data, takes a more functional view than the other two. On this view, emphasizing some aspects of experimental results and de-emphasizing others is intrinsic to the effort to produce meaningful research.

In what sense does myth result from this process of selective emphasis? One meaning of the term "myth" is "an ill-founded belief held uncritically especially by an interested group,"[2] and this is the meaning that seems to come through in Koehler's article. But if the base rate fallacy is to be considered a myth, then it is a myth in a sense similar to another meaning of the term: as a story that "serves to unfold part of the world view of a people or explain a practice, belief, or natural phenomenon."[3] In this sense, scientific myth can be seen as a device for enhancing the meaningfulness of facts, and can function like theory, which exists to be falsified and to be supplanted by better theory (Platt 1964).

Thus, while Koehler's message is worthwhile in substance, it is unnecessarily accusatory in tone. The more propitious version of Koehler's criticism is not that the base rate fallacy has been "oversold" - which makes it sound like the field has been hoodwinked - but rather, that this myth has now outlived its scientific usefulness. Interestingly, even this argument cannot be made on the basis of empirical fact alone, but it requires some theoretical

understanding of scientific myth, of how myth functions to enhance the meaningfulness of facts, and of when myth obfuscates truth rather than evincing it. Koehler's review, in revealing the base rate fallacy to be a particularly compelling myth, may provide some germinal clues in this regard (see also Koehler 1993). For example, the values and persistence of scientific myths may be related to the power of the beliefs they attempt to supplant, which, in the case of judgment fallacies, have been the stylized but captivating myths deriving from the rational theory of decision-making. Indeed, Koehler (1996, p. 43) implicitly invokes such a principle when he argues that, for his article, the "people ignore base rates" null hypothesis was a worthy target of examination because of its prevalence in the literature.

In conclusion, to more fully understand scientific belief in the base rate fallacy, it is useful to consider a functional explanation based on the observation that the meaning of an experimental result depends on the experiment's conceptual backdrop and particularly its null hypothesis. This explanation does not imply that the persistence of a base-rate-fallacy myth, or any other scientific myth, is a long-run ideal to strive for. But it does suggest that myth need not be a sign of "overselling" by researchers, and it highlights, as does Koehler's target article, the potential value of base-rate research conducted against a richer conceptual backdrop. Although, in general, there are some unique advantages to normative null hypotheses (see Kahneman 1991), Koehler is probably right in concluding that, at this point, when it comes to adding to the understanding of base rate usage, there is little left to be gained from the simple Bayesian null hypothesis. Nevertheless, it is worth acknowledging that without the conceptual structure historically provided by this normative null hypothesis, the issue of base rate usage would probably not have attracted as much research as it has, nor would it be as amenable to challenging future research of the sort envisioned by Koehler.

ACKNOWLEDGMENTS

I am grateful for valuable comments from Daniel Kahneman and Jonathan Koehler, who bear, of course, no responsibility for the opinions expressed here.

NOTES

1. More generally, this ideal of multiple perspectives may be exactly what decision makers should strive to approximate when facing a decision that can be framed in more than one way (Tversky & Kahneman 1981).
2. *Webster's Seventh New Collegiate Dictionary*, s.v. "myth."
3. Ibid.

Direct experience *is* ecologically valid

Adam S. Goodie

Center for Adaptive Behavior and Cognition,
Max Planck Institute for Human Development,
D-14195 Berlin, Germany, **goodie@mpib-berlin.mpg.de**
www.mpipf-muenchen.mpg.de/abc/people/goad-e./htm

Abstract: Koehler's (1996t) target article raised, and various commentators discussed, two issues that seem far separated but actually have a great deal in common. These are the value of "ecologically valid" research and the effect of direct experience on base-rate usage. Koehler discussed the former as a methodological issue and the latter as a normative one, and no commentator chose to incorporate them, but directly experienced base rates are a good example of ecologically valid research. The state of the literature with regard to directly experienced base rates is reviewed, and the emerging perception, that direct experience has a profound Bayesian effect on base-rate usage, is rejected.

Koehler (1996t) calls for a more ecologically valid program of research, to evaluate the extent of base-rate usage in real-world (as opposed to laboratory) situations. But what happens, once this program is undertaken, if people do not behave equally in all real-world situations? If physicians do not respond to base-rate information in the same way that lawyers do, then how can we predict the behavior of accountants? Or if physicians respond differently to base-rate information about some diseases than to base-rate information about other diseases, then how are we to predict their responses to new diseases? The answer, of course, is to identify the independent variables responsible for these differences through an experimental program, as Dawes (1996) points out. Predicting behavior in a novel case becomes a matter of evaluating the levels of the relevant independent variables.

Koehler's proposed program must proceed from the assumption that data would differ if collected in naturalistic settings. But it is almost inconceivable that naturalistic settings as diverse as a medical office, a congested freeway, and a courtroom would differ from laboratory settings in crucial ways and yet be functionally identical to each other. Koehler and several commentators try to make this claim with the property of base-rate ambiguity. But surely not every real-world situation is equally ambiguous in its base rates. If ecologically collected data differ from laboratory-data, they are almost bound to differ from each other. Once they do, the research community will not be satisfied with a report such as "accountants use base rates appropriately but lawyers do not." It will demand to know the variables responsible for the effect. The process of inquiry then reverts immediately back to the laboratory, and it should in principle be possible to manipulate any variable thought to make a difference, including base-rate ambiguity.

Experimental researchers can fruitfully speculate on the variables that might prove to be ecologically relevant. At least one such variable, the direct experiencing of base rates, has been identified and tested. It has not been given due credit for ecological validity, though, possibly because it was introduced for another purpose, namely to make the base rates more salient or vivid, not to make the entire situation more realistic. Medical students who underweight base-rate information (Casscells et al. 1978) on written tests might turn into doctors who make appropriate diagnoses. Why? It is plausible to suggest that people do not do a good job of incorporating base rates presented as summary statistics (such as on a written test), but incorporate base rates that they have experienced themselves (perhaps over many years, as a physician has after seeing many cases).

What, then, is the effect of direct experience on base-rate usage? The perception is emerging that direct experience greatly enhances Bayesian integration, but the evidence is decidedly mixed, as Fernandez-Berrocal et al. (1996) noted. Manis et al. (1980) observed substantial but incomplete base-rate integration. But Carroll and Siegler's (1977) subjects integrated base rates only when they knew that the events they predicted exhausted the class to which the stated base rates referred, and even then did so less than they should have. Gluck and Bower (1988) and Edgell et al. (1996) report having observed base-rate neglect, and Medin and Edelsen (1988) and Shanks (1992) found a tendency for subjects to predict the less likely outcome in a novel and ambiguous situation, termed the "inverse base rate effect." As Spellman (1996) indicated, I myself have observed considerable base-rate neglect following extensive and direct experience (Goodie & Fantino 1995; 1996). It should also be noted that Linderman et al. (1988), cited by Koehler as an example in this class, used a within-subjects manipulation in which each subject solved multiple verbal problems employing various base rates. Their subjects' probability estimates varied in accordance with the base rates, but when they tried to generalize similar within-subjects effects to a novel problem, they found no carryover, observing base-rate neglect even in subjects who had been informed of correct responses in previous phases. In any event, this is fundamentally different from the other studies in this class because the base rates were presented as summary statistics, not directly experienced. Even when base rates are directly experienced, the glass seems to be little more than half-full (see Keren & Thijs 1996).

Research of all kinds must always strive for ecological validity, and experimental studies stand at greater risk of sacrificing validity than field studies do. However, only experimentally demonstrated