Commentary on **Jonathan J. Koehler (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. BBS 19:1–53.**

Abstract of the original article: We have been oversold on the base rate fallacy in probabilistic judgment from an empirical, normative, and methodological standpoint. At the empirical level, a thorough examination of the base rate literature (including the famous lawyer–engineer problem) does not support the conventional wisdom that people routinely ignore base rates. Quite the contrary, the literature shows that base rates are almost always used and that their degree of use depends on task structure and representation. Specifically, base rates play a relatively larger role in tasks where base rates are implicitly learned or can be represented in frequentist terms. Base rates are also used more when they are reliable and relatively more diagnostic than available individuating information. At the normative level, the base rate fallacy should be rejected because few tasks map unambiguously into the narrow framework that is held up as the standard of good decision making. Mechanical applications of Bayes's theorem to identify performance errors are inappropriate when (1) key assumptions of the model are either unchecked or grossly violated, and (2) no attempt is made to identify the decision maker's goals, values, and task assumptions. Methodologically, the current approach is criticized for its failure to consider how the ambiguous, unreliable, and unstable base rates of the real world are and should be used. Where decision makers' assumptions and goals vary, and where performance criteria are complex, the traditional Bayesian standard is insufficient. Even where predictive accuracy is the goal in commonly defined problems, there may be situations (e.g., informationally redundant environments) in which base rates can be ignored with impunity. A more ecologically valid research program is called for. This program should emphasize the development of prescriptive theory in rich, realistic decision environments.

If the base rate fallacy is a fallacy, does it matter how frequently it is committed?

Jonathan E. Adler
Department of Philosophy,
Brooklyn College and the Graduate Center, C.U.N.Y.
Brooklyn, NY 11210. **jeabc@cunyvm.cuny.edu**

Abstract: In many base rate studies, a judgment is required for which the base rates are relevant, and subjects do not use them. It is inferred that the base rates are ignored; I question this inference. Second, I argue that the base rate fallacy is not less significant for what it reveals about human reasoning, if it occurs less frequently than has been alleged.

Although Koehler refers mainly to explicit claims that subjects ignore base rates (Koehler 1996), he does not question and seems to endorse (sect. 2, para. 1) an immediate inference from subjects' not using base rates for the task specified: that they ignore them (assigning "zero weight," in Koehler's terms). I want to deny the inference without denying that the evidence of nonuse does justify the stronger claim.

To motivate questioning that inference, consider Koehler's allusion to the argument that "base rate evidence is inconsistent with the legal norm of individualized justice and should therefore be excluded" (sect. 5.2, para. 1). Membership in a group with a known frequency of involvement with certain crimes should not serve to convict any particular person of such a crime, since it violates their right to equal treatment.

The case is not peculiar to the legal system (Adler 1994). Imagine that you teach at a high school and your class contains a clearly recognized group of mediocre students. On a recent exam you have given, the expected grades of the mediocre students are inflated. In addition, you receive an anonymous letter from someone claiming to be a student in the class, who alleges, with some telling details, that there was cheating. What do you do?

You will not flunk the mediocre students, even if you rate their probability of cheating as high. Nor will you refrain because the cost of a false accusation is more than the cost of a false grade. One solution is to make the whole class take another test, which you supervise carefully. In any event, your reasonable belief that the mediocre students have cheated is not founded on (individuating) evidence, and such evidence is required before entering an accusation. So you will attempt only corrections that impose a burden falling on all equally.

Although the base rates are not used in determining what to do, it does not follow that you ignore them. First, prior to offering another test, you might scrutinize individual papers closely for signs of cheating. The base rates might encourage you to scrutinize the mediocre students more closely than the good ones. Second, if you receive further information that confines the accusation to members of one particular group, the information may be more acceptable to you if judgment comes down with the base rates rather than if it defies them. Finally, and most obviously, the difference in base rates (students' prior grades) between the good and the mediocre students should affect your degree of belief that the cheating originated with the mediocre students.

In cases like the above, the failure to use the base rates to determine the most salient action does not imply that they are ignored. They still affect the inclination to inquire further of some individuals rather than others, one's degree of reasonable belief in such individuals, and one's willingness to modify judgment with new information.

Though the example can stand for many realistic situations, I recognize that it differs in some crucial ways from standard base rate studies. For example, we are looking to action, not a direct judgment as to how probable it is that any given student cheated. This is an "ecologically valid" or typical kind of setting where base rate information is available and known.

In such settings, the continuum of degrees of probability is not equally valuable for purposes of judgment or action. Even though full belief (or acceptance) represents only a small cluster of points on that continuum, the utility of such beliefs is much greater. Full belief, as a kind of all-or-nothing stance, provides a clearer guide to action, and it affords greater economies by ending inquiry and easing strains on memory.

It will be noted that if we assign greater weight to information that promises full belief, there are costs of error. But first, these costs, as Hogarth (1986) has suggested, are likely to be less in continuous environments, where further information is expected and belief can be modified (and thereby plans can be altered). Second, related information, such as base rate data, which may be underweighted or unused in a focal judgment, may nevertheless strongly influence one's *confidence* in that judgment (Griffin & Tversky 1992), and thereby how willing one is to modify belief with new information.

Consider the engineer-lawyer studies (Kahneman & Tversky 1973) where it appears that groups with opposed base rates give approximately the same answers. What would occur if we subsequently offered subjects a *new* item of information about some particular individual whose biography has been sampled, where the stereotype of the new information is opposed, to varying degrees, to that in the original biography? One prediction would be that the two groups would differ in their new assessment. A

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

second, more problematical prediction, is that those who judged originally in accord with the base rates would change their assessment significantly less than those who judged in opposition to the base rates.

The highly abbreviated proposal has serious problems of implementation, including whether and how the original sketch and assessment is retained, cognitive dissonance and related effects from receiving the contrary information – and, as Jonathan Baron has pointed out to me, a possible "ceiling" effect. But if not otherwise vitiated, the proposal could confirm the descriptive reality of normative models that recommend against the use of base rate information on some tasks to which they are applicable without implying that the information is discarded.

In the simplistic terms in which the debate has often been framed, my discussion so far appears to classify me with the optimists, rather than the pessimists. I have offered an example which suggests an ethnical and epistemic rationale for not using base rates on certain judgments that represent themselves as predictive and probabilistic.

Yet I do not see why, in proposing these rationales, which would justify subjects' answers, it follows that they do justify them. My doubts are sources for questions to Koehler as to the significance of his central claim that we have been oversold on the base rate fallacy.

One reason these rationales do not imply optimism is that neither one guarantees that the subject, for whom these fit behaviorally, has a corresponding understanding of these strategies. A teacher who did not use the base rates for punishing the mediocre students might think he was not entitled to a degree of belief that favored betting, on even odds, that these students were involved. (This point bears especially on sect. 2.3.3, and also, though less so, on sect. 4.2, where Koehler assumes that if subjects are semantically confused, then they cannot be committing the base rate fallacy in the same task. But why not? An erroneous view of base rates and their role in judging posterior probability might explain why subjects tend to misunderstand the key concepts they do.)

When Koehler tells us that we have been oversold, I think there is a narrower and a broader reading. The narrow one is clear enough: when one looks at the full set of base rate studies, fewer of them reveal a base rate fallacy than summary statements imply. I am not concerned with Koehler's claim on this narrow reading.

But Koehler has something broader in mind, especially with regard to the normative issues, and I find it difficult to accept this, insofar as I am able to pin it down. The idea is something like this: the base rate fallacy is not as "serious" a fallacy as we have been led to believe.

I assume that on the narrow reading the "oversold" claim is supposed to go a long distance toward yielding the broader conclusion. I can understand that a fallacy is not as significant practically if it turns out to occur less frequently than believed. But I do not accept that a fallacy is any less serious as a failure in human reasoning and understanding because it manifests itself less frequently.

Because Koehler places so much weight on the numbers, his presentation of some studies is misleading, especially that of Nisbett and his colleagues (Jepson et al. 1983) on the understanding of statistical reasoning (sects. 2.2 and 6). Koehler mentions the positive side of these studies: under various conditions, a fair percentage of subjects do reason according to valid statistical principles and these skills are teachable.

But he simply neglects the findings that motivate the educational program, and the preference for some courses (e.g., statistics) over others (e.g., chemistry, logic) of realistic problems presented to subjects. Large percentages of subjects seriously underweight the role of chance (regression, small sample size, etc.; see, for example, Jepson et al. 1983, Table 1, p. 496). The issue is not whether the glass is half full or half empty. These large percentages are the telling ones, indicative of the tendencies of our untrained reasoning.

Let me reinforce my point with another thought experiment. Imagine that we take such solid survey works as Nisbett and Ross (1980) or Piatteli-Palmarini (1994), and we cut out all references to the base rate fallacy. Base rates are extensively studied in the judgment and decision literature, and the results of those studies are used to explain others. Still, it will be granted that there is plenty of other material left, so a coherent story, with lots of supporting evidence, could still be told.

Now I ask you to think of reading these imagined books, with their discussions of difficulties with a variety of statistical principles, other probability laws (e.g., conjunction), problems in grasping randomness, and so on. In addition, these works contain a number of unifying explanatory themes (heuristics, bounded rationality, going beyond the information given, etc.) that mesh with one another.

Let us say someone asks you to predict what will occur in an equivalent to one of Tversky and Kahneman's (1973) base rate studies. The answer is, I believe, that you would predict the base rate fallacy and that you would express your prediction in existential form, and not only because of an especially cautious nature. You would not say that in all, most, or even few cases where base rates are available, they are not used. Rather, you would say that in some cases when base rates are available, subjects will not use them.

These imagined books would lead us to expect a neglect in the use of base rates, and it would be secondary how frequently this occurred. One reason is that the books have indicated that we have a tendency to downplay the influence of randomness and chance. So it would not surprise us that when the base rates are not perceived as random, they are used more. But the significant cases, even if few, would be those in which the base rates are neglected. It would be those fitting our predictions, which would be counted as ones involving less interference. These cases reveal the genuine tendency of our reasoning.

Assessing the rationality of lay social inference

Garth J. O. Fletcher

Psychology Department, University of Canterbury, Christchurch, New Zealand. **g.fletcher@psyc.canterbury.ac.nz**

Abstract: This commentary is in agreement with the thrust of Koehler's target article. The issue I deal with is whether a Bayesian framework represents an adequate general normative framework for deciding the rationality of lay judgments, even when it can be unambiguously applied.

Koehler's (1996t) target article on base rate usage is reassuringly consistent with the flood of research over the last decade or so dealing with a plethora of social inference biases, including the fundamental attribution error, the confirmatory bias, belief perseverance, and much more (see, for example, Fletcher 1995; Funder 1995; Klayman & Ha 1987). What this research shows is that, just like base rate usage, there exist a multitude of conditions under which social inference biases or errors are reduced or judgment accuracy is enhanced. In short, depending on the conditions, lay social cognition can look stupid or intelligent, scientific or unscientific.

Koehler argues that the Bayesian framework is an inadequate model of rational inference mainly because it typically fails to map unambiguously onto real-world problems (or, for that matter, onto some laboratory tasks). Several commentators questioned the normative status on the same or similar grounds. It is critically important, however, to disentangle treating the Bayesian model as a descriptive versus a normative account of lay cognition. The question I raise here is whether a Bayesian framework represents an adequate normative framework for deciding the rationality of lay judgments, even *when* it can be unambiguously applied.

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

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.stanford.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 (1996t) 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 deemphasizing 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

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

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

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

effects offer assurance of their reality, and an effect studied in a laboratory is not necessarily ecologically invalid. Some ways of studying base-rate weighting in the laboratory undoubtedly have greater ecological validity than others. I believe that directly experienced base rates lend validity to a study. Whether this is so or not, what is called for is the critical re-examination of experimental practices, not their diminution.

Base rates and randomness

Ranald R. Macdonald

Department of Psychology, University of Stirling, Stirling, FK9 4LA, Scotland. **r.r.macdonald@stir.ac.uk**
www.stir.ac/departments/humansciences/psychology/staff/rrm1

Abstract: In base rate problems the estimated probability must equal the base rate only where random sampling is assumed. Otherwise there is uncertainty over and above that which can be included in any probability model and inductive inference is involved. Subjects should use base rates to the extent that the problem suggests a simple random sampling model.

Koehler (1996) reported that, contrary to the "base rate fallacy," subjects do use base rates and they made more use of them when repeated random sampling from the base rate population was suggested by the problem. The following comments justify this practice.

Where all the relevant information for the estimation of a probability is membership of a population and its base rate then this population can contain no identifiable subsets associated with different probabilities, the event whose probability is being assessed can be regarded as having been randomly sampled and the estimated probability should equal the base rate (see Fisher 1973, p. 114). Cohen (1996), in a similar comment, referred to relevant and available information but it seems to me that the available requirement should be omitted. Randomness cannot be established in naturally occurring populations because there is no limit to the ways of classifying naturally occurring events and identifiable subsets having different probabilities may always exist. Events are members of any group which includes themselves and, providing sufficient numbers of events resulting in different outcomes exist, populations with base rates of any value between 0 and 100% can be associated with any event. Randomness has a technical and a non-technical meaning. Even when it is stated that an event has been randomly sampled, "randomly" could be interpreted as haphazardly and the event could still be described as having been selected from any number of populations each with a different base rate.

The effect a base rate has on a probability estimate should depend on how the subject conceptualizes the situation and should take account of any relevant prior knowledge, e.g., factors related to the outcomes, how similar the event in question is to other events with known outcomes, other possible base rate populations, bias in selecting the base rate, and so on (Macdonald 1986). Base rate problems involve generalizations beyond the information given and consequently require inductive inferences. They cannot be represented adequately by probability models since these are based on deductive logic and contain assumptions which are not modifiable in the light of evidence (Popper & Miller 1987). These comments apply to any statistical inference (Macdonald 1997).

As more becomes known about particular uncertainties, simple probability models may be replaced by more complex ones consisting of networks of sub-models each of which represents the uncertainty associated with events related to the event whose probability is being assessed. For instance in assessing the probability that a suspect committed a crime, one might want to assess the probability that a piece of testimony was true, which might depend on the probability that a witness was in a particular place at a particular time and that the witness was capable of seeing a particular happening. Such models quickly become very complex and Schum (1994) reported that Ward Edwards (who was responsible for establishing decisionmaking as a field of research in psychology) was working on a weather forecasting model that had 70 sub-models and required 4,500 probability assessments. In such circumstances wholly adequate models and optimal probability assessments would appear to be unobtainable idealizations.

Cohen (1996) suggested that the importance of base rates should be judged by their weight – the absolute amount of relevant knowledge and relevant ignorance used. Thus, when assessing the probability that a particular forty-year-old lorry male driver will survive to seventy, more attention should be paid to the base rate based on forty-year-old male lorry drivers than that based on forty-year-old lorry drivers. However the total amount of relevant information for assessing a particular life expectancy has no limit and so the "weight" of a base rate cannot be established. It seems to me that the importance of a base rate should be judged by the appropriateness of the probability model which assumes that the event in question has been randomly sampled from the base rate population. If alternative models seem more appropriate then the base rate is less important.

In any event an individual's uncertainty concerning an event is better represented by a number of possible probabilities and the confidence the individual has in each of them rather than by any single probability, and this representation can be further elaborated by assessing the individual's confidence in the confidences (Fisher 1957). Where random sampling (in the technical sense) is not involved, probability models are at best crude characterizations of particular states of knowledge. In making inductive inferences subjects have to generalize beyond the evidence presented to them and these generalizations should improve with an understanding of the relevant variables and their relationships. Future research might do better to concern itself with how this understanding develops rather than worrying about the accuracy of quantitative probability assessments. There is more hope for improving ways of reducing uncertainty than for establishing the true probabilities of naturally occurring events.

Pragmatically before ecologically valid tasks

Laura Macchi

Istituto di Psicologia, Universit degli Studi di Milano via Larga, 19 20122, Milan, Italy.
lmacchi@imiucca.csi.unimi.it

Abstract: Most tasks used to demonstrate the base rate fallacy are ambiguous about the independence of the data. The removal of such ambiguities from the texts (by means of a clear reference class for the probabilities) is a necessary condition and has a considerable effect on the use of the base rate in classical probabilistic tasks. Some comments are offered on the frequentist phrasing of such and, more generally, their ecological validity.

The pragmatic interpretation of the "confusion" hypothesis. Demonstrations of the role of text structure in problem solving (Mosconi 1990), and analyses of the pragmatic rules of natural language (Grice 1975; Sperber & Wilson 1986) of the texts used in an influential research paradigm (the textbook problem paradigm, Bar-Hillel 1983), suggest that the so-called base rate fallacy may arise from the fact that one relationship between the items of data is not sufficiently clear. More specifically, subjects may not perceive the two pieces of information as conditionally independent – a crucial assumption for proper Bayesian analysis (Birnbaum 1983). If subjects interpret the diagnostic information – P(D/H) – as the posterior probability – P(H/D) – (according to the "confusion" hypothesis, see para. 2.3.3 in Koehler, 1996t), they would be entitled to assume that the base rate had already been taken into account.

It has been shown (Macchi 1995) (using the two classical problems of the Cabs and the Suicide) that the base-rate fallacy

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

decreases (35%) or disappears (less than 10%) when the reference class for the relative frequency is clear, expressed as information (in terms of probability "80% of Blue cabs"; "3% of singles"), and requested as evaluation. Some textual elements (which hide the independence of the data) were modified – leaving the specificity and causality of the information unchanged. Moreover, it was possible to generate "bias" by introducing the same textual elements into "causal" versions that do not usually demonstrate the fallacy (Tversky & Kahneman 1980).

On the basis of a pragmatic analysis of the text-problems used in the literature, the differences between the problems which induce the bias and those which do not have been identified. These differences are not linked to the usual heuristic account but to the pragmatic structure of the texts.

Although this explanation is not general (other types of problems, e.g., those concerning the social judgment paradigm, Bar-Hillel 1983; Hilton & Fein 1989; Tversky & Kahneman 1980, deserve separate consideration with respect to pragmatic factors), it does explain the paradigm considered here.[1]

In this sense, it is possible to reconsider a fundamental normative and empirical point raised by Koehler about the representation of the problem. Unlike Koehler, we think that the frequentist phrasing is neither necessary nor sufficient to explain base-rate use; rather the phrasing is a particular case of clear communication of the independence of the data.

Given that base rate use has also been elicited using probabilistic phrasing tasks (Christensen-Szalanski & Beach 1982; Ginossar & Trope 1987; Macchi 1995; Manis et al. 1980), it seems reasonable to hypothesize that, regardless of the probabilistic or frequentist format of the task, the use of the base rate is linked to the effective transmission of the fundamental relationship between different pieces of information.

Furthermore, it is worth noting that a frequentist task is not sufficient in itself to elicit base-rate use. Gigerenzer and Hoffrage (1995) used only natural sampling information (in short frequency format, but also in standard frequency format, where, instead of (d/h), (d & h) [(d).(d/h)] was provided). This type of data already includes the base rate and, for this reason, should not be adopted to test the use (or neglect) or the base rate.

Only if the experiments show the use of the base rate with the information given in frequentist format corresponding to that given in classical probabilistic problems (base rate, hit rate, and false alarm rate) can frequentist phrasing be considered crucial to Bayesian reasoning, not otherwise. We have shown (Macchi 1996) that the results obtained using such a frequentist format (5 out of 100 people who are perfectly healthy) did not differ from those obtained using a probabilistic format (5% of people who are perfectly healthy). In this work, we also studied the factors that seem to be responsible for the fallacy. In both cases, once the confusion is eliminated, the results depend on other apparently secondary factors (such as computational simplicity, the explicit mention of P [not H], etc.).

Pragmatic aspects and the ecological validity of a task. We now turn to the problem raised by Koehler concerning the adequacy of paradigmatic tasks for studying the use of base rates. One question is prompted by my pragmatic analysis: which formulation is more natural, the one used in the original tasks (e.g., the original "Cab" text-problem), or the one that takes into consideration the pragmatic aspects of the communication of the problem?

It is hard to say which of these texts is more similar to the problems encountered in real life, and traditional studies do not give any clear indications in this issue. Nonetheless, we speculate that even if the first type of formulation (i.e., the original "Cab" text-problem) is more "natural," in the "real world," it is the context – the way the information is experienced – that usually removes any ambiguities concerning the relationships between the data.

In most cases, the independence of the information is still clearly transmitted (see Christensen-Szalanski & Beach's 1982 study, where there was a progressive acquisition of data in a medical context and proper use of base rates). This suggests that using a task in an experimental setting requires us to take into account the need to restore the information that has necessarily been lost by transferring it from a natural into an experimental context. Pragmatically adequate tasks can make artificial tasks ecologically valid. Artificial tasks are not necessarily artifactual ones.

NOTE

1. My point is different from the one raised by Koehler (see para. 2.2.1), since increased consideration of the base-rate was obtained using the reference class provided by the original problems (Bar-Hillel 1980; Tversky & Kahneman 1980). I consider it essential that the clarification is pragmatic concerning the information given in the text.

Nearly Bayesian uncertain reasoning methods

Paul Snow

P.O. Box 6134, Concord NH 03303-6134. **paulsnow@delphi.com**

Abstract: Subjects are reported as being somewhat Bayesian, but as violating the normative ideal on occasion. To abjure probability altogether is difficult. To use Bayes' Theorem scrupulously when weighing evidence can incur costs without corresponding benefits. The subjects' evident nuanced probabilism appears both realistic and reasonable.

Ginzburg et al. (1996) suggest that ostensibly non-Bayesian methods for reasoning under uncertainty, such as those based upon possibility theory, might be closer to the "native" uncertainty calculus of human cognition than the Bayesian paradigm. This position has a large following in the artificial intelligence default reasoning community.

An especially mature AI proposal is that of Lehmann and Magidor (1992). They define a rational default entailment relationship, styled A $|\sim$ B, whose interpretation is roughly that sentence *B* is more plausible than $\neg B$ if sentence *A* is learned to be true. Although this a kind of conditioning, it is intended to be simpler than Bayesian belief revision. Arguments for simple, unsound, but decisive default reasoning by automata are similar to those presented by Margolis (1996) in the biological context.

It is well established that all possibility measures $\Pi()$ are capable of modeling rational default entailment (Dubois & Prade 1991), using the schema A $|\sim$ B just in case that $\Pi(AB) > \Pi(A\neg B)$. Strict inequality is required, except in the degenerate case where *A* is known to be false. Only exclusive sentences are being compared.

Neither rational default entailment nor possibility theory relies on Bayesian intuition. Nevertheless, rational default entailment can also be modeled by standard subjective probabilities.

To see this, note that possibility measures are defined by the relationship $\Pi(S) = \max \Pi(s)$, where the maximum is taken over all atomic sentences *s* whose disjunction is sentence *S*. $\Pi(S) > \Pi(T)$ echoes the order of the "best supported" atoms in *S* and *T*.

Construct a system of strict *probabilistic* inequalities for each atom *s* such that $P(s) > \Sigma P(t)$, where the summation is taken over all atomic sentences *t* where $\Pi(s) > \Pi(t)$, if any. Include the usual non-negativity and total probability constraints.

The resulting simultaneous linear constraint system is consistent. A particular solution can be constructed based on the familiar properties of the powers of two. For example, one solution for five atoms is the probability distribution (1/31, 2/31, 4/31, 8/31, 16/31).

Every solution has the property that $\Pi(S) > \Pi(T)$ only if $P(S) > P(T)$ for exclusive *S* and *T*. Since only strict inequalities "count," probability distributions described by the system of constraints conform to rational default entailment, and can even use the

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

efficient "non-additive" possibilistic decision algorithm for that purpose.

More generally, *every* possibility measure represents a coherent ordering, since $\Pi(S) > \Pi(T)$ only if $P(S) > P(T)$ implies by contraposition that $P(S) >= P(T)$ only if $\Pi(S) >= \Pi(T)$. Exclusivity of *S* and *T* can be dispensed with.

A full account of the coherent orders represented by possibility measures would include a semantic treatment of possibilistic equalities and partial credal orderings, which exceeds the scope of this comment. It is worth noting, however, that the probabilistic character of possibility is not simply a matter of sterile algebra. The linear systems that tie all possibilities to some probabilities are isomorphic to systems that arise in connection with a method for nearly Bayesian and fully coherent statistical inference in the absence of prior probabilities (Snow 1996).

The moral of the default and possibility stories is that descriptions of reasonable belief revision are hard-pressed to avoid probabilistic interpretation. Even conscious aversion does not inoculate the believer against implicit probabilism. If possibility is the "native calculus," then subjective probabilistic models of that calculus exist.

Default reasoning is intentionally simple, as is typical of possibilistic applications. Consider, then, a more complicated inference task closer to routine statistical practice. Suppose evidence *E* is observed which is more likely to occur if hypothesis *H* is true than if exclusive hypothesis *G* is true. How much bigger than $P(E \mid G)$ does $P(E \mid H)$ have to be to impel belief that *H* is more credible than *G?*

Within the Bayesian paradigm, answering that question involves the prior probabilities of *H* and *G*, and hence base rates acquire importance, as noted by Klar (1996). Gigerenzer (1996) shows by an analysis of a medical screening test why prior information is useful in determining how much contrast in likelihood is enough. If *H* is rare compared to *G*, then only a very high likelihood ratio $P(E \mid H)/P(E \mid G)$ would prevent most observations of *E* being seen when *G* is true.

Of course, there are other approaches to deciding when evidentiary contrast is stark enough to induce belief. Classical statistics' tests of significance address substantially the same question using different methods and assumptions (not always scrupulously applied, as Dawes [1996] notes). In practice, the conclusions reached by classical and Bayesian statisticians are often harmonious (Edwards et al. 1963). The robustness of reliable assessment of the import of evidence is ecologically fortuitous for a species living by its wits.

Focusing on the interpretation of evidence provides perspective on the "reference class" problem discussed in so many commentaries (especially Cohen 1996, Connolly 1996, Kyburg 1996, Levi 1996, and McKenzie & Soll 1996). A quest for the elusive "actual" posterior probability is often frustrated by the multiplicity of competing populations to which any individual belongs (outside of textbook and controlled experimental problems). Nevertheless, a quick-and-dirty "thought experiment" involving a single plausible population told Gigerenzer what he really wanted to know: how bad would it be if he tested positive for HIV antibodies? In the event, he found that the test's false positive rate was swamped by the rarity of antibodies among people "like him."

These observations suggest that the marginal value of strict adherence to the single belief-revision schema of Bayes' Theorem may sometimes be slight at best. Hamm (1996) is right that his physicians would do well to consider base rates (as the workers in other domains discussed by Koonce [1996] and Quinsey [1996]), but need they do more than Gigerenzer? Could they do much more, given the conflicts about "correct" reference classes?

Koehler (1996t) reports that subjects avail themselves of prior information, using Bayes' Theorem itself more or less depending on circumstances. Several of the commentators confirm this general assessment, as noted in Koehler's (1996r) Response.

Such conclusions make sense in a world where some use of probability is apparently inevitable, but in which the Bayesian model often demands great effort and confers speculative reward beyond the insights provided by simple and robust quantitative arguments. The conditions under which subjects are found to be more or less compliant with the Bayesian ideal are consistent with the hypothesis that subjects grasp the costs and benefits of cognitive exertion in a fashion becoming to rational thinkers.

Author's Response

A farewell to normative null hypothesis testing in base rate research

Jonathan J. Koehler

Graduate School of Business and Law School, University of Texas at Austin, Austin, TX 78712. **koehler@mail.utexas. edu**
www.bus.utexas.edu/~koehlerj/

Abstract: I agree with **Gibbs** that the message of the base rate literature reads differently depending on which null hypothesis is used to frame the issue. But I argue that the normative null hypothesis, H_0: "People use base rates in a Bayesian manner," is no longer appropriate. I also challenge **Adler**'s distinction between unused and ignored base rates, and criticize **Goodie**'s reluctance to shift research attention to the field. **Macchi**'s arguments about textual ambiguities in traditional base rate problems suggest that empirical testing is needed to tease apart the effects of problem clarification and problem framing. **Macdonald**'s, **Fletcher**'s and **Snow**'s skepticism about the value of Bayesian methods in real world judgment tasks is treated as a challenge for the next generation of empirical base rate studies.

The question "What is the interesting null hypothesis in base rate research?" arose in the first round of commentaries on Koehler (1996t) and it arises again here in connection with an eloquent commentary by **Gibbs.**[1] He agrees with the substance of the target article, but criticizes what he perceives to be its harsh treatment of the normative null hypothesis. Gibbs points out that empirical data must be viewed against a conceptual backdrop. The backdrop that is commonly used in base rate research is the normative null hypothesis, H_0: "People use base rates in a Bayesian manner."[2] Gibbs suggests that the use of this null hypothesis, and the subsequent focus on the alleged base rate fallacy, "follow[s] from a reasonable judgment by scientists about what constitutes the most interesting and useful null hypothesis against which to contrast experimental results." In a paper cited by Gibbs, Kahneman (1991) provided a similar justification for using the normative null hypothesis when he argued that the "prestige" of the normative theory makes it a "sufficiently interesting" source of null hypotheses (p. 144). [See also *BBS* multiple book review of Chow: "Statistical Significance" *BBS* 21(2) 1998.]

Although there is a sense in which a normative null hypothesis is "interesting," its rejection is unlikely to be informative or otherwise useful unless it is one that reflects a reasonable expectation or widely held belief. Just as there is no longer value in testing and rejecting H_0: "The earth is flat," there is little point in rejecting the null hypothesis that subjects will give Bayesian answers in any of the classic (but impoverished) base rate problems. No one believes that people will always follow Bayes' rule in such tasks and –

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

prestige of the normative model notwithstanding – rejection of this null hypothesis is uninteresting.

In contrast, the base rate fallacy null hypothesis (H_0: "People ignore base rates") has a truth status in the literature that makes it a worthy target. Strong claims about the base rate fallacy are echoed in books, empirical studies, and review articles (see target article, sect. 1, pp. 2–3). In many cases, the earliest and most extreme data are cited to support these claims while dozens of contradictory studies are truly ignored. For example, a new book on theory and evidence repeats the conventional wisdom as follows: "[W]hen asked to make predictions, subjects ignore base-rate data and rely instead on causal theories (Tversky & Kahneman 1974; 1980)" (Koslowski 1996, p. 251).[3]

Rejection of the base rate fallacy null hypothesis is not only informative, it is needed to direct research to the more pressing issue of identifying when base rates and other forms of information are given more and less attention, and how this differential attention affects performance and satisfaction. In contrast, rejection of the normative null hypothesis is no longer helpful.

Gibbs would probably agree. Indeed, he explicitly acknowledges that "there is little left to be gained from the simple Bayesian null hypothesis." But he argues that even if the base rate fallacy myth has outlived its usefulness, this is not cause to lose faith in decision research more generally. Nor, he says, is it cause to agree with my position that we were oversold on the myth. In fact, Gibbs hints that I may be guilty of overselling *my* position.

I agree that those who accept the conclusions reached in the target article should not lose faith in behavioral decision research. But I do hope to shake the faith of those who persist in testing problems of the taxi cab variety in their ability to advance our understanding of human reasoning using such problems. There was a time when this research program was valuable and appropriate. But now it is time to replace methods that promote colorful and clever demonstrations of base rate underweighting with those that promote deeper understanding of how people think in realistic decision tasks.

As for who has oversold what, I remind the reader of what I did and did not claim. I *did* claim that the evidence fails to support the oft-heard claim that people routinely ignore useful base rate information. In doing so, I showed that there are many situations in which people attend to base rates and do so in ways that suggest a sensitivity to normative considerations (e.g., reliability and relative diagnosticity). I also emphasized the roles task structure and task representation play in determining the extent to which available base rates are used. Having disconfirmed the base rate fallacy null hypothesis (H_0: "People ignore base rates"), I did *not* claim that people routinely use base rates in accordance with Bayes' theorem.

On the other hand, many of those who have tested and rejected the normative null hypothesis (H_0: "People use base rates in a Bayesian manner") have forsaken such reasonable inferences as "There are situations in which (some) people pay relatively little attention to base rates" in favor of sweeping and misleading statements about how most people routinely "ignore" or "largely ignore" base rates.

Adler draws a distinction between base rates that are unused and base rates that are ignored. He notes that base rates that are unused for one purpose (e.g., action taken) are not necessarily unused (or ignored) for another purpose (e.g., search for additional information). This is true, but it does not mean that a distinction between "unused" and "ignored" is helpful. In the base rate literature, most people use the term "ignore" with respect to a single dependent variable. Consider Adler's example in which a high school teacher considers whether a group of students cheated on an exam. If the base rates do not affect the teacher's subjective probability estimate that the students cheated, it is appropriate to say that they were "ignored" with respect to this measure. However, if the base rates affected some other dependent variable (e.g., the teacher's willingness to monitor future tests more carefully), then the base rates have not been ignored with respect to this dependent variable. It seems to me that Adler's distinction between "unused" and "ignored" requires that we define "unused" as applicable to a single dependent measure, but define "ignored" as applicable to all possible dependent measures. I don't see why this should be the case.

Having said this, I agree with **Adler**'s implicit suggestion that it may be useful to broaden the scope of dependent measures used to study base rate usage. Adler's action-based approach fits well with my own call for more ecologically valid outcome criteria in base rate tasks.

Adler and I also agree to some extent about the overselling of the base rate fallacy. He grants that summary statements of the literature probably understate the amount of base rate usage that has been detected in the laboratory. But Adler does not believe that the base rate literature tells us less about human reasoning under uncertainty than previously believed. He writes: "I do not accept that a fallacy is any less serious as a failure in human reasoning and understanding because it manifests itself less frequently."

Adler's thought experiment – in which he asks us to predict the outcome of a base rate study based on what we know about the difficulties people have making statistical judgments from related areas of research – leads him to the following prediction: "[I]n some cases when base rates are available, subjects will not use them." Although Adler would be hard pressed to find clear support for such a prediction in the literature (see n. 3), such cases probably do exist. But if we have to work hard to produce them, if we must use unrealistic tasks and bend the definition of "not use" to mean "use, but less than that dictated by normative theory," then unqualified statements about a base rate fallacy are exaggerations. As indicated in the target article, demonstrations that base rates are ignored in certain well-defined and important real life situations would be far more valuable.

Macchi's commentary is a persuasive reminder that textual ambiguities in some classic base rate problems are partly responsible for the belief that people disregard base rates and otherwise reason in nonBayesian ways. She presents data which show that clarification of the relationships among evidentiary items in well-known base rate problems produced large increases in base rate usage. These data were cited in the target article (Koehler 1996t, sect. 2.3.3, p. 10), and were used to bolster the conclusion that inverse fallacy demonstrations (i.e., equating $P(H \mid E)$ with $P(E \mid H)$) tell us "more about the importance of effective communication of unfamiliar probabilistic notions than about the existence of an attentional or cognitive flaw" (sect. 2.3.3, p. 10).

Macchi also reports the results of a recent study to

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

support her claim that base rate usage has more to do with whether subjects understand the relationship among evidentiary items than with whether frequentist (as opposed to probabilistic) phrasing is used. The data she reports are not necessarily inconsistent with data cited in the target article which suggest that frequentist phrasing increases base rate usage. Frequentist phrasing may increase base rate usage partly because it reduces confusion about relations among evidentiary items. The lower incidence of inverse fallacy errors when data are presented as frequencies rather than as probabilities supports this contention (Thompson & Schumann 1987). Additional empirical tests that tease apart the effects of clarification from information format can help determine whether there is an effect for frequency framing beyond its value as a clarifying agent.

Goodie's point that base rates that are directly experienced are not necessarily used well by decisionmakers echoes themes expressed by Spellman (1996) and Fernandez-Berrocal et al. (1996) in the first round of commentaries. As indicated in my first response (Koehler 1996r), this is a setback for evolutionary accounts of statistical reasoning which predict that people should be well-equipped to learn base rates that appear in trial by trial learning tasks.

Goodie is also reluctant to emphasize field studies at the expense of laboratory studies. He points out that "an effect studied in a laboratory is not necessarily ecologically invalid." This is true, but why ignore the field when one of the central goals of decision research is to predict the behavior of real world decisionmakers? The issue here is one of balance. We have already studied the effects of many independent variables on base rate usage in the laboratory. But we know next to nothing about the real world conditions under which people should be advised to take more or less account of base rates.

Macdonald, Fletcher, and **Snow** comment on the limitations of the Bayesian normative model for assessing base rate usage. **Macdonald** notes that this model requires a random sampling mechanism that is rarely met in practice. He also argues that "the effect a base rate has on a probability estimate should depend on how the subject conceptualizes the situation." I agree. For example, if some subjects have goals in addition to probabilistic accuracy, then it would not be appropriate to measure subjects' performance against a single common normative standard (see target article, sect. 5.2, p. 15). This is not to say that either Macdonald or I believe that any response a subject provides in a base rate task is appropriate. But it does suggest that our standards for what constitutes good judgment should be tailored more closely to the assumptions and goals of the individual decisionmaker.

Similarly, **Fletcher** notes that, even when mapping problems such as those discussed in section 4 of the target article (p. 12–13) are not a serious concern, adherence to Bayesian norms is neither a necessary nor sufficient condition for assessing rationality. Because Bayes' theorem fails to (a) provide guidance for identifying values for prior probabilities and the diagnosticity of information, and (b) incorporate such meta-values as the desire for a stable belief system, it provides little insight into the rationality of human decisionmakers.

I agree with the strong form of **Fletcher**'s argument, but caution that nonBayesian performance may signal shortcomings in how people think about evidence and diagnosticity. This may reduce performance on important outcome measures (e.g., business earnings, legal verdicts, medical recommendations). Of course, these are empirical issues. But if future study indicates that real world performance is improved through adherence to Bayesian principles, it would seem misguided to remain silent on grounds that Bayes is an incomplete authority on rational inference.

Snow points out that there are alternatives to the Bayesian belief updating model. He also notes the ecological fortuity of the fact that conclusions reached by the alternatives are often similar (see also McKenzie 1994). These observations support my contention that the natural environment may well be structured such that the penalties for failure to adhere to Bayesian norms often are mild (see target article, sect. 5.1, pp. 14–15). Likewise, Snow's claim that reference class ambiguities limit the value of base rate information in real life decision tasks suggests that those who routinely use base rates may not outperform those who do not. Such empirical propositions should be tested in the next generation of base rate research.

NOTES

1. Far less eloquent is a related point I made in my response to the first round of commentaries (Koehler 1996r). The first full sentence that appeared on page 43 was mangled beyond comprehension during the typesetting process. It should have read: "Whereas few researchers suggested that people process base rates in a Bayesian fashion, many suggested that people attach no weight to base rates (cites)."

2. This discussion holds aside the difficulties associated with assuming base rates are identical to subjects' Bayesian priors (see target article, sect. 4.1, pp. 12–13).

3. Similarly, it is interesting to note that thoughtful commentators continue to repeat the conventional wisdom that subjects in the lawyer-engineer studies "give approximately the same answers" regardless of the base rate (**Adler**). Although the means and medians of low and high base rate groups are often close, Table 1 of the target article shows that the high base rate groups that included diagnostic individuating information yielded higher means or medians than the low base rate groups in all 15 lawyer-engineer experiments (exact binomial, $p < .0001$). Consequently, a claim that responses in high and low base rate groups are approximately the same is akin to a claim that men and women are approximately the same height, or that the number 1 and number 2 seeds in tennis tournaments have approximately the same chance of winning.

References

Adler, J. E. (1994) More on race and crime. *Journal of Social Philosophy* 25:105–14. [JEA]

Bar-Hillel, M. (1980) The base-rate fallacy in probability judgements. *Acta Psychologica* 44:211–33. [LM]

(1983) The base-rate fallacy controversy. In: *Decision-making under uncertainty*, ed. R. W. Scholtz. Elsevier. [LM]

Birnbaum, M. H. (1983) Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology* 96:85–94. [LM]

Carroll, J. S. & Siegler, R. S. (1977) Strategies for the use of base-rate information. *Organizational Behavior and Human Performance* 19:392–402. [ASG]

Casscells, W., Schoenberger, A. & Graboys, T. B. (1978) Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine* 299:999–1001. [ASG]

Christensen-Szalanski, J. J. J. & Beach, L. R. (1982) Experience and the base-rate fallacy. *Organizational Behavior and Human Performance* 29:270–78. [LM]

Cohen, L. J. (1996) The need for a theory of evidential weight. *Behavioral and Brain Sciences* 19(1):18–19. [RRM, PS]

Connolly, T. Are base rates a natural category of information? *Behavioral and Brain Sciences* 19(1):19–20. [PS]

Dawes, R. M. (1996) The purpose of experiments: Ecological validity versus comparing hypotheses. *Behavioral and Brain Sciences* 19(1):20–21. [ASG, PS]

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press

Dubois, D. & Prade, H. (1991) Possibilistic logic, preferential models, non-monotonicity, and related issues. *Proceedings of the International Joint Conference on Artificial Intelligence* 419–24. [PS]
Edgell, S. E., Roe, R. M. & Dodd, C. H. (1996) Base rates, experience, and the big picture. *Behavioral and Brain Sciences* 19(1):21. [ASG]
Edwards, W., Lindman, H. & Savage, L. J. (1963) Bayesian statistical inference for psychological research. *Psychological Review* 70:193–242. [PS]
Fernandez-Berrocal, P., Almaraz, J. & Segura, S. (1996) How to reconsider the base rate fallacy without forgetting the concept of systematic processing. *Behavioral and Brain Sciences* 19(1):21–22. [ASG, rJJK]
Fisher, R. A. (1957) The underworld of probability. *Sankhya* 18:201–10. [RRM]
(1973) *Statistical methods and scientific inference*. 3rd. edition. Hafner. [RRM]
Fletcher, G. J. O. (1995) *The scientific credibility of folk psychology*. Erlbaum. [GF]
(1996) Realism versus relativism in psychology. *American Journal of Psychology* 109:409–29. [GF]
Funder, D. C. (1995) On the accuracy of personality judgement: A realistic approach. *Psychological Review* 102:652–70. [GF]
Gigerenzer, G. (1996) Why do frequency formats improve Bayesian reasoning? Cognitive algorithms work on information, which needs representation. *Behavioral and Brain Sciences* 19(1):23–24. [PS]
Gigerenzer, G., Hell, W. & Blank, H. (1988) Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perceptions and Performance* 14(3):513–25. [LM]
Gigerenzer, G. & Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102(4):684–704. [LM]
Ginossar, Z. & Trope, Y. (1987) Problem solving in judgement under uncertainty. *Journal of Personality and Social Psychology* 52:464–74. [LM]
Ginzburg, L. R., Janson, C. & Ferson, S. (1996) Judgment under uncertainty: Evolution may not favor a probabilistic calculus. *Behavioral and Brain Sciences* 19(1):24–25. [PS]
Gluck, M. A. & Bower, G. H. (1988) From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General* 117:227–47. [ASG]
Goodie, A. S. & Fantino, E. (1995) An experientially derived base-rate error in humans. *Psychological Science* 6:101–06. [ASG]
(1996) Learning to commit or avoid the base-rate error. *Nature* 380:247–49. [ASG]
Grice, H. P. (1975) Logic and conversation. In: *Syntax and semantics: Vol. 3. Speech Acts*, ed. P. Cole & J. L. Morgan. Academic Press. [LM]
Griffin, D. & Tversky, A. (1992) The weighing of evidence and the determination of confidence. *Cognitive Psychology* 24:411–35. [JEA]
Hamm, R. M. (1996) Physicians neglect base rates, and it matters. *Behavioral and Brain Sciences* 19(1):25. [PS]
Hilton, J. L. & Fein, S. (1989) The role of diagnosticity in stereotype-based judgements. *Journal of Personality and Social Psychology* 57:201–11. [LM]
Hogarth, R. M. (1986) Beyond discrete biases. In: *Judgement and decision making*, ed. H. R. Arkes & K. R. Hammond. Cambridge University Press. [JEA]
Jepson, C., Krantz, D. H. & Nisbett, R. E. (1988) Inductive reasoning: Competence or skill. *Behavioral and Brain Sciences* 3:494–501. [JEA]
Kahneman, D. (1991) Judgement and decision making: A personal view. *Psychological Science* 2:142–45. [BJG, rJJK]
Kahneman, D. & Tversky, A. (1973) On the psychology of prediction. *Psychological Review* 80:237–51. [JEA]
Keren, G. & Thijs, L. J. (1996) The base rate controversy: Is the glass half-full or half-empty? *Behavioral and Brain Sciences* 19(1):26. [BJG, ASG]
Klar, Y. (1996) P(D/H), P(D/H) and base rate consideration. *Behavioral and Brain Sciences* 19(1):26. [PS]
Klayman, J. & Ha, Y. W. (1987) Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review* 94:211–28. [GF]
Koehler, J. J. (1993) The influence of prior beliefs on scientific judgements of evidence quality. *Organizational Behavior and Human Decision Processes* 56:28–55. [BJG]
(1996t) The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences* 19(1):1–17. [JEA, GF, BJG, ASG, rJJK, LM, RRM, PS]
(1996r) Issues for the next generation of base rate research. *Behavioral and Brain Sciences* 19(1):41–53. [rJJK, PS]
Koonce, L.(1996) Base rates in the applied domain of accounting. *Behavioral and Brain Sciences* 19(1):29. [PS]
Koslowski, B. (1996) *Theory and evidence*. MIT Press. [rJJK]
Kuhn, T. (1970) *The structure of scientific revolutions (2nd ed.)*. University of Chicago Press. [BJG]
Kyburg, H. E., Jr. (1996) Probabilistic fallacies. *Behavioral and Brain Sciences* 19(1):31. [PS]
Lehmann, D. & Magidor, M. (1992) What does a conditional knowledge base entail? *Artificial Intelligence* 55:1–60. [PS]
Levi, I.(1996) Fallacy and controversy about base rates. *Behavioral and Brain Sciences* 19(1):31. [PS]
Lindeman, S. T., Van Den Brink, W. P. & Hoogstraten, J. (1988) Effect of feedback on base-rate utilization. *Perceptual and Motor Skills* 67:343–50. [ASG]
Macchi, L. (1995) Pragmatic aspects of the base-rate fallacy. *The Quarterly Journal of Experimental Psychology* 48A(1):188–207. [LM]
(1996) Biases in probabilistic reasoning: Heuristics, frequentist phrasing and pragmatica factors. In: *Perspectives and limits of cognitive economics, Quaderni Lascomes, n. 1*, ed. R. Viale. Universit Bocconi. [LM]
Macdonald, R. R. (1986) Credible conceptions and implausible probabilities. *British Journal of Mathematical and Statistical Psychology* 39:15–27. [RRM]
(1997) On statistical testing in psychology. *British Journal of Psychology* 88:333–47. [RRM]
Manis, M., Dovalina, I., Avis, N. E. & Cardoze, S. (1980) Base rates can affect individual predictions. *Journal of Personality and Social Psychology* 38:231–48. [ASG, LM]
Margolis, H. (1996) Nuancing should not imply neglecting. *Behavioral and Brain Sciences* 19(1):32. [PS]
McKenzie, C. R. M. (1994) The accuracy of intuitive judgement strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology* 26:209–39. [rJJK]
McKenzie, C. R. M. & Soll, J. B. (1996) Which reference class is evoked? *Behavioral and Brain Sciences* 19(1):34–35. [PS]
Medin, D. L. & Edelsen, S. M. (1988) Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General* 117:63–85. [ASG]
Mosconi, G. (1990) *Discorso e pensiero*. Il Mulino. [LM]
Nisbett, R. & Ross, L. (1980) *Human inference: Strategies and shortcomings of social judgement*. Prentice Hall. [JEA]
Peter, J. P. & Olson, J. C. (1983) Is science marketing? *Journal of Marketing* 47:111–25. [BJG]
Piatteli-Palmarini, M. (1994) *Inevitable illusions*. John Wiley & Sons. [JEA]
Platt, J. R. (1964) Strong inference. *Science* 146:347–53. [BJG]
Popper, K. & Miller, D. W. (1987) Why probabilistic support is not inductive. *Philosophical Transactions of the Royal Society, London* A 321:569–91. [RRM]
Quinsey, V. L. Improving decision accuracy where base rates matter: The prediction of violent recidivism. *Behavioral and Brain Sciences* 19(1):37. [PS]
Schum, D. A. (1994) *Evidential foundations of probabilistic reasoning*. Wiley. [RRM]
Shanks, D. R. (1992) Connectionist accounts of the inverse base-rate effect in categorization. *Connection Science* 4:3–18. [ASG]
Snow, P. (1996) Inference using conditional probabilities despite prior ignorance. *IEEE Transactions on Systems, Man, and Cybernetics* 26-A:349–60. [PS]
Spellman, B. A. (1996) The implicit use of base rates in experiential and ecologically valid tasks. *Behavioral and Brain Sciences* 19(1):38. [ASG, rJJK]
Sperber, D. & Wilson, D. (1986) *Relevance: Communication and cognition*. Basil Blackwell. [LM]
Thompson, W. C. & Schumann, E. L. (1987) Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior* 11:167–87. [rJJK]
Thomsen, C. J. & Borgida, E. (1996) Throwing out the baby with the bath water? Let's not overstate the overselling of the base rate fallacy. *Behavioral and Brain Sciences* 19:39–40. [BJG]
Tversky, A. & Kahneman, D. (1974) Judgements under uncertainty: Heuristics and biases. *Science* 185:1124–30. [rJJK]
(1980) Causal schemas in judgements under certainty. In: *Progress in social psychology, vol. 1.*, ed. M. Fishbein. Erlbaum. [rJJK, LM]
(1981) The framing of decisions and the psychology of choice. *Science* 211:453–58. [BJG]

https://doi.org/10.1017/S0140525X97251642 Published online by Cambridge University Press