Second, accounting tasks frequently lack an outcome referent that can be used to gauge decision quality (Davis & Solomon 1989). For example, when judging the inherent risk of an audit, or the reliability of a company's system of internal controls, there really is no correct answer. As a consequence of this absence of a correctness measure in accounting, researchers have turned to other measures, such as judgment/decision defensibility (Peecher & Kleinmuntz 1991). Such a focus away from "accuracy" is consistent with the basic premise of Koehler's analysis. That is, he suggests that we relax our notion of what constitutes an appropriate response to accommodate the unique characteristics of different judgment/decision contexts.

Conclusion. Koehler successfully challenges the base rate fallacy. His claim that base rates are used is consistent with the findings from accounting research. Moreover, his recommendation that researchers should investigate base rate usage in realistic decision contexts with flexible standards of acceptance performance is sound – particularly for areas such as accounting, which lack outcome referents to judge decision quality.

Studying the use of base rates: Normal science or shifting paradigm?

Joachim Krueger

Walter S. Hunter Laboratory of Psychology, Brown University, Providence, RI 02912. **joachim-krueger@brown.edu**

> Normal-scientific research is directed to the articulation of those phenomena and theories that the paradigm already supplies.
>
> Kuhn (1962/1970, p. 24)

Abstract: The underutilization of base rates is a consistent finding. The strong claim that base rates are ignored has been rejected and this needs no further emphasis. Following the path of "normal science," research examines the conditions predicting changes in the degree of underutilization. A scientific revolution that might dethrone the heuristics and biases paradigm is not in sight.

The fascinations and frustrations arising from research on people's use of base rates follow a Kuhnian course. Initially, the discovery of systematic flaws in human inference met with skepticism, which was subsequently replaced by enthusiasm and careful study. As in other fields, the strong early claims were gradually moderated and clarified. In this process, the meaning of the base rate fallacy changed from "complete neglect" to "insufficient use." The systematic exploration of the variation of the size of this effect exemplifies what Kuhn (1962/1970) meant by normal science. In contrast, Koehler suggests that there is no good evidence for a pervasive base rate fallacy, and his argument seems to imply that a revolutionary shift from a Bayesian to an ecological paradigm is in order. The rejection of the strong claim of complete base rate neglect is like breaking down open doors. As Koehler himself notes, "some (but not all) commentators have softened their views" (intro., para. 5). The more modest claim, namely, that the use of base rates is insufficient, is well supported and in no need of a new paradigm.

In section 1, Koehler cites numerous studies that, as he concedes, document insufficient use of base rates. He cites few studies that indicate strong use of base rates. Two of these studies do not permit tests of relative over- or underuse (Biernat 1993; Nelson et al. 1990). Others, although they do show base rate effects, also support the modest claim of insufficient use. Wells and Harvey (1977, p. 286) found that subjects "underutilized" base rates when diagnostic information was present. Hilton and Fein (1989), according to Koehler, found "robust effects of stereotypes on trait attributions" (sect. 1.2, para. 2), but this is not surprising, because all individuating information was irrelevant (see also Hamilton 1984). Quoting from Hilton and Fein's abstract, Koehler states that "only individuating information that is 'useful across many social judgment tasks' can reduce the impact of stereotype base rates" (sect. 1.2, para. 2). What this means is that even nondiagnostic information, when made to look pseudorelevant, can reduce the effect of base rates.

Our study (Krueger & Rothbart 1988) also supported this modest claim. We tested the idea that stereotypic beliefs about social groups, for better or for worse, influence inferences about individual group members. When such beliefs are probabilistic, Bayes' rule specifies optimal inferences. Suppose the perceived base rate of trait T (e.g., aggressive) is $p(T) = .4$ for a man and .1 for a woman. Without relevant individuating information, inferences about individuals depend solely on the base rate. Now suppose that a certain behavior (e.g., yelling) is diagnostic of aggressiveness. If a person is aggressive, yelling has the probability of $p(B|T) = .7$; if a person is not aggressive, not yelling has the probability of $p(-B|-T) = .7$. Note that the probability of the behavior, p(B), differs for men and women because of the difference in the base rates. For men, $p(B) = p(T)p(B|T) + p(-T)p(B|-T) = .4 \times .7 + .6 \times .3 = .46$. For women, $p(B) = .1 \times .7 + .9 \times .3 = .34$ (see also target article, n. 2). It follows from these assumptions that the probability that a man who yells is aggressive is .61 (i.e., $.4 \times .7/.46$), and the probability that a woman who yells is aggressive is .21 (i.e., $.1 \times .7/.34$).

As this example shows, two things happen when base rates differ between groups and when the behavioral information is equally diagnostic of the trait for both groups. First, the posterior probabilities for members of both groups go up. Consistent with this prediction, men and women who had yelled at someone were rated to be more likely to be aggressive than men and women who had not yelled. Second, because Bayes' rule is multiplicative, the between-group differences become larger. Posterior probabilities do not converge but *diverge*, given diagnostic information. The sex difference in the hypothetical example increases from .3 to .4. Contrary to this prediction, our subjects seemed to combine group stereotypes and individuating information *additively*. The sex difference did not increase with the introduction of diagnostic behavior. These data suggest that subjects neither ignore base rates, as some have claimed (e.g., Locksley et al. 1980), nor do they weigh them sufficiently.

One study, according to Koehler, showed overutilization of base rates. Rasinski et al. (1985) found that subjects underestimated the probability that a target person is assertive, given that person's assertive behavior and given the subjects' own estimates of base rates and behavior diagnosticity. What the data also showed, however, was that subjects *ignored the prior sex differences*. The optimal difference between predictions about men and women was +7%, given the diagnostic information (Table 1, p. 319). Subjects' ratings, however, showed a difference of −3% (Table 2, p. 320). Violating Bayes' rule, subjects judged men to be slightly less likely to be assertive than women.

Koehler's review demolishes the already defunct claim that people completely ignore base rates. The modest claim of base rate underutilization, however, remains viable and full of possibilities. Judgmental biases need not be absolute. To my knowledge, no bias reported in the literature satisfies the totalitarian criterion. Following the path of normal science, researchers have articulated the boundary conditions and the contextual variations of the base rate fallacy and other biases, and will continue to do so. Koehler's "ecologically valid research program" contributes important ideas for future directions. The program seems to expand rather than overturn the current heuristics and biases paradigm.