ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 42, 343–363 (1988)

Pseudodiagnosticity in Judgment under Uncertainty

CHEZY OFIR

Hebrew University of Jerusalem

The study investigates the extent to which the false alarm (i.e., $P(D/\bar{H})$) is utilized in judgment under uncertainty. The main findings are (1) this cue is utilized by subjects when provided with a numerically low base-rate (i.e., $P(H)$) and a high hit-rate (i.e., $P(D/H)$). Under these conditions the false alarm helps resolve the inconsistency between the implications of these probabilistic cues. (2) The false alarm is ignored by subjects when provided with a numerically high base-rate and a high hit-rate. Under these conditions both latter probabilities are consistent and imply strong support for the focal hypothesis. The false alarm is either not consistent with these cues or redundant and ignored. In addition, three experiments provided evidence regarding base-rate utilization. When comparing conditions (1) and (2) it is demonstrated that the base-rate has a significant effect on judgments. Finally, the experiments provided evidence suggesting that the base-rate is not ignored when the datum is not diagnostic. The results are discussed in relation to the base-rate fallacy, pseudodiagnosticity, and cue consistency.

In evaluating whether a hypothesis (H) is true relative to its complement ($\bar{H}$) one should take into consideration, according to the Bayesian model, prior odds ($P(H)/P(\bar{H})$) and the likelihood ratio. This ratio indicates the informativeness of the evidence (D) with regard to the truth of the hypothesis (H). The numerator of this ratio ($P(D/H)$) is an indication of the probability of the datum given the focal hypothesis—the hit-rate. The denominator ($P(D/\bar{H})$) indicates the probability of the datum (D) given the alternative hypothesis ($\bar{H}$)—the false alarm.

Much of the research dealing with human judgment under uncertainty in a Bayesian framework has focused on the neglect or underutilization of base-rate information. Judgments are dominated by the information extracted from the evidence while prior beliefs are ignored or largely underutilized. This bias is known as the base-rate fallacy (for reviews see Borgida & Brekke, 1981; Kassin, 1979; Sherman & Corty, 1984; Tversky & Kahneman, 1982). A recent critical review of research in this domain concluded that yet another "metabias" may exist, namely "the tendency

The author acknowledges the support of the Shey Foundation, the Krueger Center, and the Faculty of Social Sciences, Hebrew University. He also acknowledges helpful comments of David Budescu, John Lynch, Yehoshua Tsal, and Yaacov Trope on a previous draft of the paper. Address for correspondence: School of Business Administration, Hebrew University, Mount Scopus, Jerusalem, 91905, Israel.

0749-5978/88 $3.00

to ignore $P(D/\bar{H})$ when evaluating evidence" (Fischhoff & Beyth-Marom, 1983, p. 257). It was further suggested that people are mainly interested in the consistency of the evidence with the hypothesis under consideration (i.e., $P(D/H)$) and tend to ignore its consistency with the alternative hypothesis (i.e., $P(D/\bar{H})$).

Some evidence bearing on this bias was furnished by Doherty, Mynatt, Tweney, and Schiavo (1979). Most subjects in their experiment requested probabilities relating to their favored hypothesis, while only very few subjects requested the two probabilistic cues needed to assess the diagnosticity of each piece of datum according to Bayes' theorem. This bias was termed pseudodiagnosticity. Beyth-Marom and Fischhoff (1983) pointed out, however, that it is not clear whether subjects in the Doherty *et al.* study perceived no value in the probability related to the alternative hypothesis or just assigned to it insufficient weight relative to $P(D/H)$.

On the basis of five experiments, Beyth-Marom and Fischhoff (1983) concluded that, in general, people are much better at using the false alarm (i.e., $P(D/\bar{H})$) than seeking it out or providing the reasons for its usage. In particular, in their experiment focusing on cue utilization (Experiment 5, pp. 1191–1193), they demonstrated that subjects use the probability related to the alternative hypothesis. They pointed out, however, that subjects were provided with data which were either not diagnostic at all or with data which were highly diagnostic. The question arises, as noted by Beyth-Marom and Fischhoff (1983), as to what happens when both components of the likelihood ratio support their respective hypotheses (e.g., $P(D/H) = 90\%$ and $P(D/\bar{H}) = 60\%$).

Conditions for False Alarm Utilization

Based on these empirical results it would seem that people use the false alarm in some conditions. These conditions, however, are presently not specified. Some insight may be gained by examining the role of cue consistency and the base-rate in these judgments. Consider, for example, the Cab problem (e.g., Fischhoff & Bar-Hillel 1984a; Lyon & Slovic, 1976; Tversky & Kahneman, 1982) which reads as follows:

> Two cab companies operate in a given city, the Blue and the Green (according to the color of the cab they run). Ten percent of the cabs in the city are Blue, and the remaining 90% are Green.
>
> A cab was involved in a hit and run accident at night. A witness later identified the cab as a Green cab.
>
> The court tested the witness' ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found that the witness was able to identify Green Cabs correctly 90% of the time. The witness, however, confused it with the

> other color 60% of the time. Namely, when the cab was Blue the witness indicated that it was Green 60% of the time.
>
> What do you think is the probability (expressed as a percentage) that the cab involved in this accident was indeed Green?

In this version of the problem, the base-rate (i.e., 90% Green cabs in the city) and the hit-rate of the witness in identifying Green cabs (i.e., P(witness says green/Green) = 90%) are consistent, implying support for the focal hypothesis (i.e., Green cab). The implication of the false alarm (i.e., P(witness says green/Blue) = 60%), however, is inconsistent with these cues.

Previous research on cue utilization suggests that when cues have inconsistent implications, subjects focus on one of the cues and use other cues to resolve the inconsistency (Slovic, 1966). In this version of the Cab problem, subjects are hypothesized to look at the high hit-rate (i.e., P(witness says green/Green)) and the base-rate (i.e., P(Green)) which are directly related to the hypothesis under consideration (Beyth-Marom & Fischhoff, 1983), and to turn to the false alarm (i.e., P(witness says green/Blue)) when these latter cues are not consistent.

Consistency is used here in relation to the implication of each cue considered separately. In the above example, a witness with a high hit-rate says that it is a Green cab. The base-rate implies a Green cab as well, since most cabs in the city are Green. These two cues, therefore, have consistent implications. In contrast, the false alarm is rather high, implying a Blue cab. Given the highly consistent implication of the former two cues, this latter cue is hypothesized to be either ignored or to have a minor effect on judgments.

A somewhat different situation may arise when the false alarm is numerically lower than 50%. In particular, very low values (e.g., P(witness says green/Blue = 10%)) imply, indirectly, support for the focal hypothesis. In other words, the witness indicates green when the cab is Blue only infrequently. In this case, this cue could be either ignored on grounds that it is perceived as not adding information and highly redundant (Slovic & Lichtenstein, 1971) *or* used with equal weights to the other consistent cues (Slovic, 1966). Given the tendency of people to focus on probabilistic information pertaining to the focal hypothesis (i.e., P(D/H)) and the high consistency of the base-rate with this cue, the above hypothesis applies here as well.

Summarizing, given a high hit-rate and a high base-rate which have consistent implications supporting the focal hypothesis (i.e., H-Green cab), it is hypothesized that the false alarm will be ignored or have only a minor effect on judgments regardless of its numerical level.

Consider now the same problem in a city with 10% Green cabs and 90% Blue cabs. Here the base-rate and the hit-rate have inconsistent implica-

tions, but the base-rate and false alarm have consistent implications. The base-rate (i.e., P(Green) = 10%) implies that most cabs in the city are Blue, and the false alarm suggests that in the test when the cab was Blue, the witness indicated incorrectly Green cab 60% of the time. Both cues suggest some support for the alternative hypothesis, i.e., a Blue cab.

More generally, given a high hit-rate (e.g., P(witness says green/Green) = 90%), a low base-rate (e.g., P(Green) = 10%) and a false alarm (i.e., P(witness says green/Blue)) greater than 50%, the hit-rate stands in contrast to the two other probabilistic cues. Since the hit-rate (i.e., P(D/H)) is related directly to the focal hypothesis, it will most likely be utilized by subjects (Beyth-Marom & Fischhoff, 1983). The other two cues, however, will operate to reduce its implications. Under these conditions it is hypothesized that the false alarm (i.e., $P(D/\bar{H})$ – P(witness says green/Blue)) will not be ignored.

Given the same conditions for the base-rate and the hit-rate (i.e., P(Green) = 10%, and P(witness says green/Green) = 90%), a false alarm (i.e., P(witness says green/Blue)) less than 50% implies a Green cab (i.e., H). Here the false alarm is consistent with the hit-rate, and again it is hypothesized that it will not be ignored.

The above discussion suggests that the false alarm operates to help resolve the inconsistency between the implications of the low base-rate and the high hit-rate. When false alarm is above 50% it is consistent with the base-rate. When it is below 50% it is consistent with the hit-rate. It is thus expected that it will not be ignored at all probability levels. It is also hypothesized that given these conditions, probability judgments (i.e., P(H/D) – P(Green/witness says green)) will decrease as a function of the false alarm.

The consistency of the false alarm (i.e., P(witness says green/Blue)) with either the low base-rate or with the high hit-rate is strongest when it takes values at either end of the probability scale. It would seem that just above and below these probability levels this consistency substantially weakens. A false alarm in the neighborhood of zero indicates that the witness very rarely says green when the cab is Blue. This supports the implication of the high hit-rate (i.e., P(witness says green/Green) = 90%) both implying a Green cab. As the false alarm increases this consistency weakens. Since the false alarm will not provide as strong a support for the implication of the hit-rate, the effect of the low base-rate (i.e., P(Green) = 10%) will increase. In this region (i.e., 10–30%) a drop in the probability judgment is expected. Similar logic suggests a major change in judgments immediately below the upper end of the probability scale (i.e., 70–90%). In the middle range (i.e., 30–70%), judgments as a function of the false alarm are expected to decrease but with a moderate slope. In this region

the support of the false alarm to either the implications of the base-rate or the hit-rate is much less powerful than at the extremities of the probability scale.

Conditions for Base-Rate Utilization

The above discussion implicitly specifies also a condition for base-rate utilization. A strong effect of the base-rate is expected when the datum is very likely under both the focal hypothesis (e.g., P(witness says green/Green) = 90%) and the alternative hypothesis (e.g., P(witness says green/Blue) = 80%). In particular, given a high hit-rate, a numerically high base-rate will be consistent with this latter probability operating to reinforce a high probability judgment (i.e., P(Green/witness says green)). A low base-rate, however, will be consistent with a false alarm larger than 50% both implying a Blue cab, operating to lower the probability judgments. When the false alarm is less than 50%, it is consistent with the high hit-rate implying a Green cab. This consistency is a function of the numerical level of the false alarm. In other words, when the false alarm (i.e., P(witness says green/Blue)) is close to zero its consistency with the high hit-rate (i.e., P(witness says green/Green)) is very high suggesting that the low base-rate (i.e., P(Green)) will have minimal effect. As the false alarm increases, the consistency with the hit-rate weakens and the effect of the base-rate increases.

Summarizing, given a high hit-rate, varying the values of the base-rate will have an effect on judgments. Moreover, a base-rate by false alarm interaction is predicted; i.e., the effect of the base-rate will be minimal when the false alarm is very low. This effect is expected to increase as a function of the level of the false alarm.

The following experiments were designed to test these hypotheses and to address some of the research questions raised by Beyth-Marom and Fischhoff (1983).

EXPERIMENT 1

The base-rate and false alarm were the variables manipulated in this between subjects experiment. Given a numerically low base-rate and a high hit-rate, the normative Bayesian solution (i.e., P(Green/witness says green)) drops rapidly as a function of false alarm (see Fig. 1). If indeed people ignore or largely underutilize the false alarm, we would expect a flat curve. If, however, people utilize this cue, judgments are expected to be significantly different from a control group in which false alarm is not provided. Moreover, if people use the false alarm in the way hypothesized above, the pattern of judgments will decrease and fit more closely the normative, nonlinear curve.

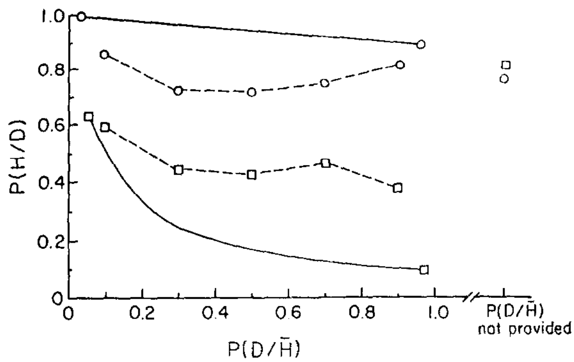

FIG. 1. Normative and subjective judgment as a function of the false alarm and the base-rate. ○———○, base = 0.90, normative solution; □———□, base = 0.10, normative solution; ○— — —○, base = 0.90, experimental results; □— — —□, base = 0.10, experimental results; ○ —, base = 0.90, experimental results; □ —, base = 0.10, experimental results.

Under the condition of a numerically high base-rate and a high hit-rate, the normative solution will change moderately as a function of false alarm. The curve obtained by the normative solution is presented in Fig. 1. The hypothesis developed for this condition implies a flat curve since the false alarm is assumed to be either ignored or to have only a minor effect on judgment.

Method

Stimuli. The cab problem, presented above, was employed in this experiment.

Design. False-alarm (i.e., P(witness says green/Blue)) and the proportion of Green cabs in the city were manipulated factorially in a 2 × 6 completely randomized between subjects design. The base-rate (i.e., P(Green)) levels were 10% and 90%, and the values of the false alarm were 10%, 30%, 50%, 70%, 90%, and a control group in which this cue was not provided. The hit-rate (i.e., P(witness says green/Green)) was fixed in all 12 cells at 90%.

Subjects. Subjects were 240 undergraduate Social Science students who were randomly assigned to the 12 cells resulting in 20 subjects per cell.

Results

An ANOVA was performed on judgments from the 2 × 5 experimental subdesign without the two control groups.[1] Both cues had a significant

[1] The mean square error used in this test is taken from the complete design with 12 cells.

impact on judgments though the base-rate effect was larger. Significant effects of the base-rate ($F(1,228) = 80.42$, $p < .0001$) and marginally significant effect of the false alarm ($F(4,228) = 2.29$, $p < .06$) were found. The interaction in this analysis was not significant ($F(4,228) < 1$).

The means of judgments in the various cells and the curves implied by the normative Bayesian model are presented in Fig. 1. Nonorthogonal planned comparisons between mean judgments and the appropriate control groups were conducted.[2] As hypothesized, when the base-rate was low (i.e., 10%) all five tests were significant, indicating utilization of the false alarm. When the base-rate was high (i.e., 90%) none of the tests were significant indicating that subjects probably did not use the false alarm.

Finally, the utilization of the base-rate is affected by the presence of the false alarm. In particular, the simple effect of the base-rate in each level of false alarm is significantly different from its effect in the absence of this cue.[3] This can be seen in Fig. 1 by comparing the vertical distance between mean judgments of the control groups to the vertical distance between mean judgments in each level of the false alarm.

Discussion

Findings regarding false alarm utilization. As expected, when the base-rate was low subjects did not ignore the false alarm as evident by the comparisons to the control group. They even used it in cases where both components of the likelihood ratio supported their respective hypotheses. Moreover, examination of Fig. 1 reveals that the pattern of judgments somewhat resembles the nonlinear normative curve. It seems that people combine the cues in a way that produces judgments in the right direction but not quite as prescribed by the normative rule. The expected interaction, however, which was implied by the discussion regarding cue consistency, was not significant. The pattern of judgments, when the base-

All other tests presented in the paper involving subdesigns and contrasts are also computed using the mean square error of the complete designs, respectively.

[2] These and all other planned contrasts in this experiment were conducted with the Sidak (1967; see also Kirk, 1982) procedure. The following are the results for the first condition (i.e., the base-rate equals 10%) respectively: $F(1,228) = 10.17$ ($p < .005$), $F(1,228) = 24.8$ ($p < .001$), $F(1,228) = 27.46$ ($p < .001$), $F(1,228) = 22.47$ ($p < .001$), $F(1,228) = 33.87$ ($p < .001$). When the base-rate was 90%, none of the tests are significant ($F(1,228) = 1.35$, $F(1,228) < 1$, $F(1,228) < 1$, $F(1,228) < 1$, $F(1,228) < 1$).

[3] The following are the results of contrast which compares the simple effect of the base-rate with its effect without $P(D/\bar{H})$, respectively: $F(1,228) = 9.46$ ($p < .005$), $F(1,228) = 9.77$ ($p < .005$), $F(1,128) = 10.58$ ($p < .005$), $F(1,228) = 11.04$ ($p < .001$), $F(1,228) = 21.62$ ($p < .001$).

rate is low, decreases moderately in comparison to the hypothesized pattern. It would seem that mean judgment based on a low base-rate (P(Green) = 10%), a low false alarm (P(witness says green/Blue) = 10%), and a high hit-rate is somewhat lower than expected (mean = 58.7). Under this condition the hit-rate and the false alarm are highly consistent. The low base-rate is inconsistent with these cues and its effect was expected to be minimal. Indeed, the median which is less sensitive to discrepant responses (e.g., Fischhoff & Bar-Hillel, 1984) was much higher (median = 80) suggesting the existence of an interaction. This issue will be further tested in the following experiments.

As hypothesized, the effect of the false alarm seems to diminish when the base-rate is high. In this situation judgments are probably not affected by any numerical level of this cue. In particular, means of judgments in all experimental cells are not significantly different from the mean of a control group in which the false alarm was not provided.

Results obtained in some of the conditions here are comparable to those of previous research. In particular, given high base-rate and hit-rate, most subjects in Beyth-Marom and Fischhoff's Experiment 5 provided judgments either higher or equal to their initial judgment. This pattern was evident given either a low (93.7%) or high (79.2%) false alarm and is consistent with findings presented here.

Findings regarding base-rate utilization. The base-rate in this experiment was utilized and had a larger effect on judgments than did false alarm. Judgments associated with a high base-rate (mean = 77.4) were significantly higher than those associated with a lower base-rate (mean = 46.2) as demonstrated by the two separate curves in Fig. 1.

Given nondiagnostic evidence (i.e., P(witness says green/Green) = P(witness says green/Blue) = 90%) the base-rate has an impact on judgments. Specifically, given a base-rate of 10% and a likelihood ratio of one, mean judgment was 38.2, median 35, and a normative Bayesian solution of 10%. Given a base-rate of 90%, mean judgment was 82.3, median 89.5, and a normative solution of 90%. These results seem to be consistent with Beyth-Marom and Fischhoff's (1983, Experiment 5) findings. About half the subjects in each of four experimental cells with nondiagnostic data (i.e., likelihood ratio equals one) did not change their initial assessment and continued to rely mainly on the base-rate. Overall, the results of the experiment presented here suggest that the base-rate was utilized.

The pattern of judgments based on a low base-rate and a high hit-rate as a function of false alarm decrease only moderately in comparison to the expected pattern. These conditions will be further examined in the following experiments. These one-way experiments will also enable replication of findings of false alarm utilization.

EXPERIMENTS 2 AND 3

Method

Stimuli. The problem used in Experiment 2 was the Disease problem previously employed to assess underutilization of the base-rate (Hammerton, 1973; Lyon & Slovic, 1976). It reads as follows:

1. A device has been invented to test for Disease X.
2. The device is a very good one, but not perfect.
3. If someone suffers from X, there is a 90% chance that he will be recorded positively.
4. If he is not a sufferer, there is still a 10% chance that he will be recorded positively.
5. Roughly 10% of the population has the disease.
6. Mr. Smith has been tested and the result is positive. The chance that he is in fact a sufferer is ——%.

The problem used in Experiment 3 was the Cab problem (see above).

Design. The design in Experiment 2 was a completely randomized one-way between subjects design. The false alarm (i.e., P(positive result/not a sufferer)) was manipulated in 10 experimental cells starting with zero with equal increments of 10% up to 90%.[4] The 11th cell was a control group in which subjects did not receive this cue. The prevalence of the disease in the population (i.e., the base-rate) and the hit-rate (i.e., P(positive result/sufferer)) were fixed in all 11 cells at 10% and 90%, respectively.

Experiment 3 consisted of a completely randomized one-way between subjects design. The base-rate (i.e., P(Green)) and the hit-rate (P(witness says green/Green)) were fixed at 15% and 80%, respectively. The false alarm (i.e., P(witness says green/Blue)) was manipulated with the following levels: 20%, 40%, 60%, and 80%.

Subjects. Subjects in Experiment 2 were 242 American students at the Rothenberg School for Overseas Students of the Hebrew University who were recruited individually and randomly assigned to the 11 cells resulting in 22 subjects per cell.

In Experiment 3, 96 undergraduate students from the same population (i.e., Experiment 2) were randomly assigned to four cells resulting in 24 subjects per cell.

Results and Discussion

Experiment 2. The medians of judgments in the various cells are plotted in Fig. 2 along with the curve implied by the Bayesian model. The judgments, as expected, decrease as a function of the false alarm. Comparing median judgments in most of the conditions to the median of the control

[4] The stimuli in the first cell indicated that the false alarm of the device is "very small . . . close to zero (much less than one tenth of one percent)."

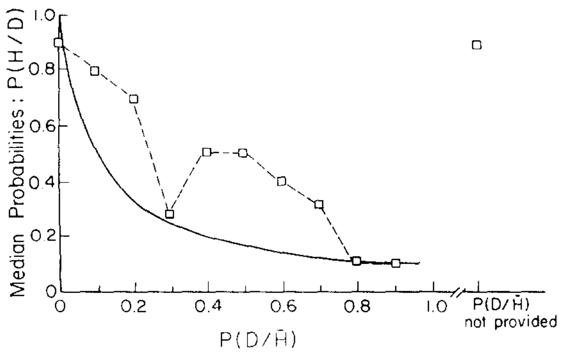

FIG. 2. Normative and subjective judgments as a function of the false alarm. ———, base = 0.10, normative solution; □— — —□, base = 0.10, experimental results; □, base = 0.10, experimental results.

group suggested that providing the false alarm does have an effect on judgments.

In a more formal analysis of mean judgments, a one-way ANOVA was conducted with the subdesign containing the 10 experimental cells. A significant main effect was obtained for the value of the false alarm ($F(9,231) = 8.25$, $p < .001$) suggesting that it had a significant impact on judgments. Further examination of the pattern of mean judgments suggests that the effect was mainly concentrated in a linear and a quartic trend components.[5] However, the pattern of medians (see Fig. 2), which are less sensitive to discrepant judgments, seems to be in line with the hypothesized pattern with a moderate slope in the middle region and sharper slopes at the extremes of the scale. An unexplainable exception to this shape of the function are judgments provided when the false alarm was 30%.

The false alarm was utilized in all experimental cells except when it equaled 0%, since all the contrasts between each experimental cell and the control group were significant.[6]

Experiment 3. The medians of judgments in the various cells are plotted in Fig. 3. In this experiment as well, judgments decrease as a function of

[5] These trend contrasts and the additional nonorthogonal planned contrasts were executed with the Sidak (1967) procedure. The following are the results for the orthogonal polynomial trend components: Linear $F(1,231) = 53.64$ ($p < .001$), Quadratic $F(1,231) < 1$ (n.s.), Cubic $F(1,231) = 8.32$ ($p < .005$, but not significant according to the Sidak procedure), Quartic $F(1,231) = 51.32$ ($p < .001$).

[6] The following are the results of the comparisons to the control group respectively (all tests are with 1 and 231 degrees of freedom): $F = 2.69$ (n.s.), $F = 9.89$ ($p < .003$), $F = 12.46$ ($p < .001$), $F = 40.2$ ($p < .001$), $F = 22.75$ ($p < .001$), $F = 13.25$ ($p < .001$), $F = 38.44$ ($p < .001$), $F = 28.41$ ($p < .001$), $F = 45.97$ ($p < .001$), $F = 71.06$ ($p < .001$).

the false alarm. The results of a one-way ANOVA suggest that, in general, judgments are sensitive to variation in the false alarm ($F(3,92) = 7.63$, $p < .002$). Inspection of Fig. 3 also suggests that, as expected, the slope of the judgment curve was steeper at both extremes of the range than in the middle. A comparison of the slope of the judgment curve in the middle section (i.e., false alarm: 40%–60%) versus the average slopes of the other two sections of the curve (i.e., false alarm: 20%–40%, and 60%–80%) lead to the same conclusion ($F(1,92) = 6.45$, $p < .025$).

Findings regarding base-rate utilization. In both experiments, when the evidence was not diagnostic, subjects seemed to be responsive to the base-rate. In Experiment 2 they provided judgments with a mean of 19.05, median 10, while the normative solution is 10%. In Experiment 3, given a base-rate of 15% and both components of the likelihood ratio 80%, mean judgment was 25.3, median 20, and a normative result of 15%. Consistent with Experiment 1, the results of Experiments 2 and 3 suggest that the base-rate was probably utilized. Here, however, the one-way design does not permit a firm conclusion regarding base-rate utilization.

Mean judgments presented in Experiment 1 suggest that judgments based on a low base-rate and a high hit-rate decrease only moderately as a function of the false alarm. This was not in accordance with the hypothesized pattern. Examination of the medians, however, revealed a pattern more in line with expectations. Judgments in these conditions were further investigated in Experiments 2 and 3: The results, as expected, indicated a decreasing function. In the following experiment, the false alarm and the base-rate were manipulated while the hit-rate was fixed at a numerically high level in all the experimental conditions. In this experiment, which is similar to Experiment 1, the pattern of judgments given low and high base-rates will be further examined and an interaction is expected.

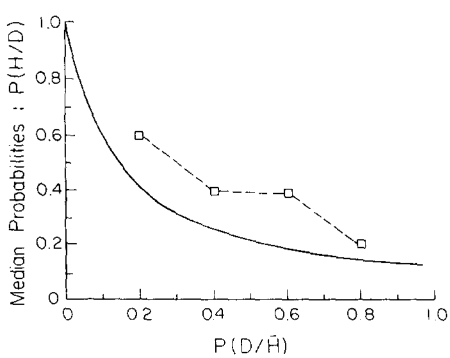

FIG. 3. Normative and subjective judgments as a function of the false alarm. ———, base = 0.15, normative solution; □— — —□, base = 0.15, experimental results.

EXPERIMENT 4

Method

Stimuli. The problem used in this experiment is a modified Disease problem. It reads as follows:

> A device has been invented to test for a disease known as diabetes. The device is now available in pharmacies and drug stores. Mr. Smith has been tested and the result is positive.
>
> The following information has been provided by the manufacturer (and approved by the appropriate governmental agency) regarding the reliability of the device and the prevalence of the disease:
>
> 1. If someone has the disease, there is a 90% chance that the test will provide a positive result (a positive result indicates that the patient has the disease).
> 2. If he does not have the disease, there is a 10% chance that the test will provide a positive result.
> 3. Roughly 5% of the population group of the same ethnic origin and age as Mr. Smith has the disease.
>
> What is the probability that Mr. Smith is in fact a sufferer? ——%

Design. The base-rate and the false-alarm were manipulated in a 2×3 completely randomized between subjects design. The base-rate levels were 5% and 90%, and the values of the false alarm were 10%, 50%, and 90%. The hit-rate was fixed in all six experimental cells at 90%.

Subjects. Subjects were 120 undergraduate American students at the Rothenberg School for Overseas Students at the Hebrew University who were recruited individually and randomly assigned to the six cells resulting in 20 subjects per cell.

Results and Discussion

Median probability judgments are presented in Fig. 4. Given a high base-rate (see upper curve in Fig. 4) the pattern of judgments is, as expected, flat suggesting that the false alarm was either ignored or had only minor effects on judgments. As hypothesized, given a low base-rate, judgments decrease, indicating that the false alarm was not ignored and the existence of an interaction.

The analysis applied to the corresponding means provided similar results. An ANOVA was performed on judgments from the 2×3 design revealing significant effects of the base-rate ($F(1,114) = 27.42, p < .001$) and the false alarm ($F(2,114) = 5.03, p < .01$). The interaction in this analysis was significant ($F(2,114) = 3.56, p < .03$).

As hypothesized, the false alarm had no effect on judgments when the base-rate was high ($F(2,114) < 1$) but had a significant effect on judgments when the base-rate was low ($F(2,114) = 8.26, p < .01$).

Examining the effect of the base-rate at different levels of the false

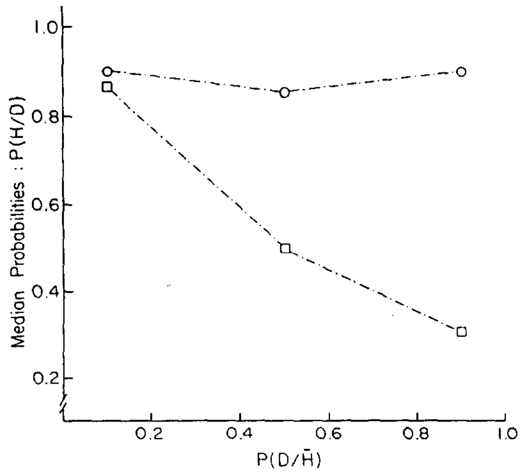

FIG. 4. Probabilistic judgments as a function of the base-rate and the false alarm. ○—·—○, base = 0.90; □—·—□, base = 0.05.

alarm suggests that, as hypothesized, given a low false alarm, the base-rate had no significant effect ($F(1,114) = 1.1$). For higher values of false alarm, however, the base-rate had significant effects (for false alarm of 50% and 90% respectively: $F(1,114) = 10.05$, $p < .01$, and $F(1,114) = 23.38$, $p < .001$).[7]

The latter simple effect is of special interest since the likelihood ratio in these conditions equals 1 (i.e., P(D/H) = P(D/H̄) = 90%). Consistent with previous results (i.e., Experiments 1, 2, and 3), it is clear that the base-rate has an effect on judgments. Given a base-rate of 5%, mean judgment was 37.1, median 30.5, and a normative result of 5%. When given a base-rate of 90%, mean judgment was 79.3, median 90, and a normative result of 90%.

In the following experiment, a replication of the findings regarding utilization of the false alarm and the base-rate given with a high hit-rate are provided. In addition, in this experiment the utilization of these cues will be examined given a low hit-rate.

EXPERIMENT 5

In this experiment the base-rate, hit-rate, and false alarm were factorially manipulated. When the base-rate and false alarm are given with a high hit-rate, their utilization pattern is expected to be in line with the hypotheses and findings presented above. This experiment also provides the opportunity to examine their utilization in the presence of a low hit-

[7] Simple effects of both cues were tested in this experiment following Dunn's procedure (Kirk, 1982, p. 369).

rate. A low hit-rate (e.g., P(witness say green/Green) = 20%)) implies a Blue cab and the pattern of judgments is expected to change. In line with the discussion regarding cue consistency, the following are hypotheses developed for conditions with a low hit-rate.

Hypotheses regarding judgments based on a low hit-rate, a high base-rate, as a function of the false alarm. In these conditions the high base-rate (e.g., P(Green) = 85%) implies a Green cab, and the low hit-rate (e.g., P(witness says green/Green) = 20%) implies a Blue cab. The false alarm plays a tie-breaking role in these conditions and is not hypothesized to be ignored. In particular, a low false alarm (e.g., P(witness says green/Blue) = 20%) implies like the high base-rate (e.g., P(Green) = 85%) a Green cab, while the low hit-rate implies a Blue cab. Under these conditions high judgments are expected. A high false alarm (e.g., P(witness says green/Blue) = 90%), however, implies like the low hit-rate, a Blue cab. These two cues operate to lower the implication of the high base-rate (e.g., P(Green) = 85%) which implies the focal hypothesis (i.e., a Green cab). In this latter condition judgments are hypothesized to be lower than in the previous condition (i.e., a low hit-rate, a low false alarm, and a high base-rate). The pattern of judgments given a high base-rate and a low hit-rate, therefore, is expected to decrease as a function of the false alarm.

Hypotheses regarding judgment based on a low hit-rate, a low base-rate, as a function of the false alarm. In these conditions both the low hit-rate (e.g., P(witness says green/Green) = 20%) and the low base-rate (e.g., P(Green) = 15%) imply a Blue cab. A high false alarm (e.g., P(witness says green/Blue) = 80%) also implies a Blue cab. Under these conditions the latter cue is expected to be ignored since it does not add information to the former two highly consistent cues and therefore is redundant. Judgments (i.e., P(Green/witness says green)) in this condition are hypothesized to be very low.

A low false alarm (e.g., P(witness says green/Blue) = 20%), however, implies a Green cab. This implication is inconsistent with that of the low hit-rate (e.g., P(witness says green/Green) = 20%) and the low base-rate (e.g., P(Green) = 15%), indicating a Blue cab. The false alarm in this condition is the only cue which supports the focal hypothesis (i.e., a Green Cab) and is hypothesized to be utilized. Since the former two cues suggest a Blue cab and a low probabilistic judgment, the false alarm, indicating a Green cab, will operate to increase the judgment to some extent.

The pattern of judgments based on a low hit-rate as a function of the false alarm is hypothesized to be similar and to decrease given either a high or a low base-rate. The above discussion, however, suggests that

judgments will be lower with a low base-rate. A base-rate by false alarm interaction, given a low hit-rate, is therefore not expected.

Since an interaction of base-rate by false alarm is expected given a high hit-rate, and not hypothesized given a low hit-rate, a triple interaction is expected in the following three-factor experiment.

Method

Stimuli. The Cab problem, presented above, was employed in this experiment. The only change in the text is related to the false alarm which was shortened: "It also found that when the cab was Blue the witness indicated that it was Green 20% of the time."

Design. The base-rate, hit-rate, and false alarm were manipulated factorially in a 2 × 2 × 3 completely randomized between subjects design: The base-rate (i.e., P(Green)) levels were 15% and 85%, the values of the hit rate (i.e., P(witness says green/Green)) were 20% and 80%, and the values of the false alarm (i.e., P(witness says green/Blue)) were 20%, 80%, and a control group in which this cue was not provided.

Subjects. Subjects were 240 undergraduate students who were recruited individually and randomly assigned to the 12 conditions resulting in 20 subjects per cell.

Results

The medians of the judgments in the various conditions are presented in Fig. 5. As hypothesized, an interaction of base-rate by false alarm given a high hit-rate is indicated by the pattern of judgments (see left panel of

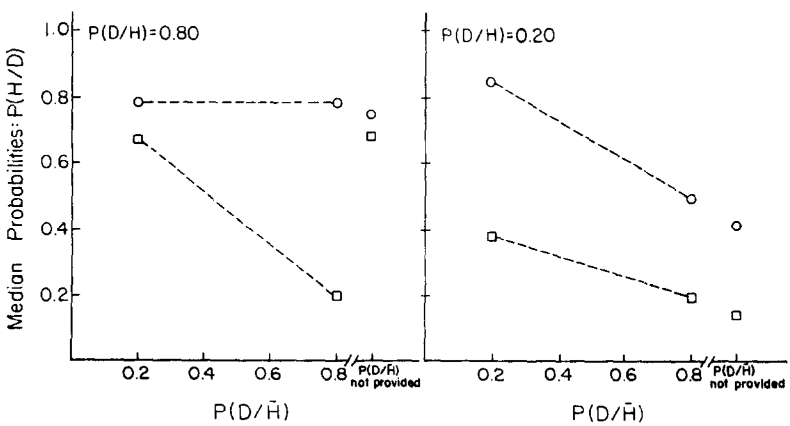

FIG. 5. Judgments as a function of the hit-rate, false alarm, and the base-rate. ○— — —○, base = 0.85, experimental results; □— — —□, base = 0.15, experimental results; ○—, base = 0.85, experimental results; □—, base = 0.15, experimental results.

Fig. 5). Such an interaction does not seem to be indicated, however, given a low hit-rate (see right panel of Fig. 5). In general, the pattern of median judgments in the various conditions is consistent with the expected patterns: Judgments are not responsive to the false alarm given a high hit-rate and a high base-rate as evident by the flat curve and the similar median judgment in the control group. Judgments, however, decrease as a function of the false alarm given a high hit-rate and a low base-rate. Inspection of the pattern of judgments based on a low hit-rate (see right panel of Fig. 5) suggests that, as hypothesized, judgments decrease as a function of the false alarm. Judgments, as expected, are higher with a high base-rate. In these latter conditions, the slope is slightly steeper than with a lower base-rate. Overall, a similar pattern of results is revealed by the means of the various conditions. The following is a formal analysis applied to these means.

Judgments based on the base-rate, hit-rate, and false alarm. An ANOVA was performed on mean judgments from the $2 \times 2 \times 2$ subdesign without the control groups. Each of the three cues had a significant effect on judgments (base-rate, $F(1,228) = 95.69$, $p < .0001$; false alarm, $F(1,228) = 32.36$, $p < .0001$; and the hit-rate, $F(1,228) = 21.71$, $p < .0001$). As hypothesized, the triple interaction was significant ($F(1,228) = 4.52$, $p < .05$). This interaction is ordinal to the effects of all three cues. The locus of the interaction is evident in Fig. 5; the interaction base-rate by false alarm given a low hit-rate is not significant ($F(1,228) < 1$). This interaction, however, is significant in the presence of a high hit-rate ($F(1,228) = 4.05$, $p < .05$). In this latter condition (see left panel of Fig. 5) the false-alarm has no effect in the presence of a high base-rate ($F(1,228) = 1.51$), but has a significant effect given a low base-rate ($F(1,228) = 16.63$, $p < .001$).

Comparisons of judgments based on the base-rate, hit-rate, and false alarm to judgments in the control groups. Consistent with previous results, the false alarm is not utilized given a high hit-rate and a high base-rate as evidenced by the nonsignificant comparisons of mean judgments with the control group (for a low false alarm, $P(D/\bar{H}) = 20\%$, $F(1,228) < 1$, and for a high false alarm $P(D/\bar{H}) = 80\%$, $F(1,228) < 1$). When the base-rate is low and the hit-rate high, a low $P(D/\bar{H})$ is ignored ($F(1,228) = 3.26$, n.s.), but a high false alarm is utilized ($F(1,228) = 34.61$, $p < .001$).

Similar contrasts were conducted to test the utilization of the false alarm given with a low hit-rate (see right panel of Fig. 5). As expected, a low false alarm is utilized given either with a high base-rate ($F(1,228) = 15.55$, $p < .001$), or with a low base-rate ($F(1,228) = 10.44$, $p < .005$). Also in line with the corresponding hypothesis, a high false alarm is ignored when provided with a low base-rate and a low hit-rate ($F(1,228) < 1$). However, not as hypothesized, a high false alarm is ignored when

provided with a high base-rate ($F(1,228) < 1$). In this condition the high false alarm apparently does not add much information to the low hit-rate since both imply a Blue cab.

The effect of the value of the false alarm on the utilization of the base-rate. The value of the false alarm may also have had an effect on base-rate utilization. Given a high false alarm (i.e., P(witness says green/Blue) = 80%) and a high hit-rate, the impact of the base-rate on judgments as expressed by the simple effect, is different from its simple effect in the absence of the false alarm ($F(1,228) = 13.37, p < .001$). This can be seen by comparing the vertical distance between mean judgments of the control groups (to the right of the left panel of Fig. 5) to the vertical distance of mean judgments when P(witness says green/Blue) = 80%. It seems that the base-rate has no effect on judgments in the absence of the false alarm, but has a major effect when a high false alarm is provided. The simple effect of the base-rate, however, is not significantly different when provided either with, or without a low false alarm = 20% ($F(1,228) = 2.69$).

Identical contrasts were executed in the low hit-rate conditions (see right panel of Fig. 5). Here the presence of any level of the false alarm does not affect the simple effects of the base-rate (for P(witness say green/Blue) = 20%, $F(1,228) < 1$, and for (P(Witness says green/Blue) = 80%, $F(1,228) < 1$).

Base-rate utilization given nondiagnostic evidence. As in previous experiments, it is of interest to examine base-rate utilization in the presence of nondiagnostic evidence (i.e., P(witness says green/Green) = P(witness says green/Blue)). When both components of the likelihood ratio are 80% and a base-rate of 15% was provided, mean judgment was 31.0, median 20, and a normative result of 15%. Given a base-rate of 85%, mean judgment was 69.5, median 77.5, and a normative result of 85%.

Similarly, in conditions where both components of the likelihood ratio equaled 20% and given a base-rate of 85%, mean judgment was 71.85, median 85.0, and a normative result of 85%. When a base-rate of 15% was provided, mean judgment was 35.15, median 37.5, and a normative result of 15%.

Discussion

In general, it is evident that the false alarm is not ignored. This conclusion is based on the significant main effect of this cue, on comparison of judgments in the various conditions with the control groups, and on the fact that the value of false alarm affects the utilization of the base-rate.

The pattern of false alarm utilization given a high hit-rate is in line with the "consistency" hypotheses and results presented in Experiments 2–4. Moreover, the preliminary findings for most conditions with a low hit-rate are also consistent with hypotheses developed on similar grounds.

Finally, judgments based on a high base-rate are significantly higher (mean = 66.7) than judgments based on a lower base rate (mean = 35.8). Overall, it is evident that the base-rate had a large effect on judgments. Moreover, its utilization pattern is predicted by the consistency hypotheses.

GENERAL DISCUSSION

False Alarm Utilization

The cumulative results support the hypothesis that people do not ignore the false alarm which is provided with a high hit-rate and a low base-rate. The false alarm role in these conditions is to help resolve the inconsistency between these latter cues. A false alarm above 50% implies, like the low base-rate (e.g., P(Green) = 10%), the alternative hypothesis (e.g., a Blue cab). A false alarm less than 50% implies, like the high hit-rate (e.g., P(witness says green/Green) = 90%), the focal hypothesis (e.g., a Green cab).

The only exception to these findings is that the false alarm is ignored when it takes values close to zero. With very low values the false alarm provides strong support for the implication of the hit-rate (i.e., H). Under these conditions the base-rate has either a minor effect on judgments or is ignored since it stands in sharp contrast to the two other probabilistic cues, which support the focal hypothesis. Under such conditions, the hit-rate seems to dominate judgments. The judgments may have been almost identical to those in the control condition since both the false alarm and the hit-rate imply high judgments.

Overall, under the conditions of a low numerical level of the base-rate and a high hit-rate, people do not ignore the false alarm and are responsive to changes in the numerical levels of the cue. This would appear to suggest that people are probably sensitive to the diagnosticity of the evidence as indicated by Trope and Bassok (1982) in the context of social information gathering.

In contrast, when a high base-rate and a high hit-rate are provided, there is no evidence that subjects use the false alarm to any significant degree. In this situation, the base-rate and the hit-rate are highly consistent, implying strong support for the hypothesis under consideration. The false alarm, which is either inconsistent with these cues ($P(D/\overline{H}) > 50\%$) or redundant ($P(D/\overline{H}) < 50\%$), is either ignored or has only a minor effect on judgments.[8]

Interestingly, the pattern of judgments given a high base-rate is somewhat similar to that obtained by the normative model. Since subjects

[8] It should be noted that a powerful test to trace small differences would require a very large sample size.

probably ignored the false alarm as evident by comparisons to the control group, and thus were not responsive to the diagnosticity of the evidence, it is obvious that their judgments were based on a rationale inconsistent with the logic of the normative model. In other words, in some conditions they provided fairly close to optimal judgments for the "wrong" reasons.

Base-Rate Utilization

The cumulative results indicate that base-rate information was utilized. Judgments based on nondiagnostic evidence (e.g., $P(D/H) = P(D/\bar{H}) = 90\%$) are very responsive to this latter cue. In fact, judgments are close to the normative result provided by Bayes' theorem. These results are in line with the hypotheses that a high base-rate is consistent with a high hit-rate operating to reinforce a high probability judgment, while a low base-rate is consistent with the high false alarm, both probabilities operating to lower the implication of the high hit-rate.

These results, however, seem to be inconsistent with those of Bar-Hillel (1980; see for a description of her results, Fischhoff & Bar-Hillel, 1984a) who reported that the base-rate was ignored when both components of the likelihood ratio equaled 50%. Under this condition, in line with the logic of the above hypotheses, both components of the likelihood ratio are consistent and may imply complete uncertainty. Either a low or a high base-rate is inconsistent with these cues and therefore may be either ignored or discounted.

More generally, as hypothesized, judgments are very responsive to variation in the numerical levels of the base-rate given a high hit-rate. In Experiments 1, 4, and 5, the base-rate had a significant effect on judgments. The results of Experiment 5 also suggest that judgments are sensitive to variation in the numerical levels of the base-rate given a low hit-rate (see right panel of Fig. 5). Overall it is evident that the base-rate was utilized.

Previous systematic findings which were inconsistent with the base-rate fallacy were based on (1) changing the content of the problem in order to manipulate the causality (Ajzen, 1977; Tversky & Kahneman, 1980) and specificity (Bar-Hillel, 1980) of the base-rate; (2) changing the problem in order to reduce the representativeness of the evidence so that it is not useful in predicting category membership (Fischhoff & Bar-Hillel, 1984b; Ginosar & Trope, 1980); and (3) changes in the procedure and/or format of the task (e.g., Christensen-Szalanski & Beach, 1982; Fischhoff, Slovic, & Lichtenstein, 1979; Manis, Dovalina, Avis, & Cardoze, 1980).

In contrast, here, the effect of the base-rate was demonstrated without either manipulating any surface details of the problem which could make this cue appear more causal and specific, or changing the procedure and format. In fact, the problems and experimental procedures employed here

are identical to those used to demonstrate the base-rate fallacy. The numerical levels of the cues, however, were differently designed to create conditions of cue consistency as described above.

The base-rate fallacy, however, was found to exist in only one of the conditions examined in this research. Judgments are not very responsive to changes in the numerical levels of the base-rate given with a high hit-rate and a low false alarm. This underutilization of the base-rate, however, is predicted by the "consistency" hypotheses. Specifically, in this condition, the high hit-rate (e.g., P(witness says green/Green) = 90%) and the low false alarm (e.g., P(witness says green/Blue) = 10%) imply quite strongly the focal hypothesis (i.e., H – Green cab); the base-rate (e.g., P(Green) = 10%) implies the alternative hypothesis (i.e., H – Blue Cab) and is therefore discounted.

Do People Possess a Latent Understanding of Diagnosticity?

The problem people have in seeking out the false alarm (Beyth-Marom & Fischhoff, 1983), other biases found in judgment under uncertainty, and the results presented here, suggest that people do not possess a latent understanding of the Bayesian model. Alternatively, Beyth-Marom and Fischhoff (1983) suggested that people may not be as bad at finding out intuitively what to do with numerical probabilistic information, and then the relevant question is "How deep that spontaneous understanding goes?" (Beyth-Marom & Fischhoff, 1983, p. 1194). The research presented here suggests that the apparent understanding demonstrated by subjects is driven by judgmental strategies based on cue consistency. Some of the limits of this understanding could be examined by comparing judgments with curves implied by the Bayesian model. The pattern of judgments is not unlike the Bayesian curves but not quite as prescribed by this model. Given a low base-rate, and a high hit-rate, judgments drop as a function of the false alarm but do not drop as low as implied by the normative model.[9] Given high base-rate and hit-rate, the pattern of judgments is also somewhat similar to the normative curve. In this latter case, it would seem that subjects provided judgments without any consideration of the diagnosticity of the evidence and thus without any latent understanding of the concept of diagnosticity.

REFERENCES

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology,* **35,** 303–314.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica,* **44,** 211–233.

[9] The relationship between the stated probability and the subjective underlying variable is assumed to be an identity function.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology,* 45(6), 1185–1195.

Borgida E., & Brekke, N. (1981). The base-rate fallacy in attribution and prediction. In J. H. Harvey, W. J. Ickes, & R. H. Kidd (Eds.), *New directions in attribution research* (Vol. 3). Hillsdale, NJ: Erlbaum.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1982). Experience the base-rate fallacy. *Organizational Behavior and Human Performance,* **29**, 270–278.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica,* **43**, 111–121.

Fischhoff, B., & Bar-Hillel, M. (1984a). Focusing techniques: A shortcut to improving probability judgments? *Organizational Behavior and Human Performance,* **34**, 175–194.

Fischhoff, B., & Bar-Hillel, M. (1984b). Diagnosticity and base-rate effect. *Memory and Cognition,* **12**(4), 402–410.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review,* **90**, 239–260.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance,* **23**, 339–359.

Ginosar, Z., & Trope, Y. (1980). The effects of base-rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology,* **16**, 228–242.

Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology,* **101**, 252–254.

Kassin, S. M. (1979). Consensus information, prediction, and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology,* **37**, 1966–1981.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences.* Belmont, CA: Brooks/Cole Publishing Co.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base-rates in probability estimation. *Acta Psychologica,* **40**, 287–298.

Manis, M., Dovalina, I., Avis, N. E., & Cardoze, S. (1980). Base-rates can affect individual predictions. *Journal of Personality and Social Psychology,* **38**, 231–248.

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer and T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 189–286). Hillsdale, NJ: Erlbaum.

Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association,* **62**, 626–633.

Slovic, P. (1966). Cue consistency and cue utilization in judgment. *American Journal of Psychology,* **79**, 427–434.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance,* **6**, 649–744.

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social-information gathering. *Journal of Personality and Social Psychology,* **43**, 22–34.

Tversky, A., & Kahneman, D. (1980). Causal schemes in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1982). Evidential impact on base-rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty—Heuristics and biases.* New York: Cambridge Univ. Press.

RECEIVED: August 25, 1986