Acta Psychologica 44 (1980) 211–233

THE BASE-RATE FALLACY IN PROBABILITY JUDGMENTS

Maya BAR-HILLEL*

Hebrew University, Jerusalem and *Decision Research, A Branch of Perceptronics, Inc., Eugene, OR*

Revised version received February 1979

The base-rate fallacy is people's tendency to ignore base rates in favor of, *e.g.*, individuating information (when such is available), rather than integrate the two. This tendency has important implications for understanding judgment phenomena in many clinical, legal, and social-psychological settings. An explanation of this phenomenon is offered, according to which people order information by its perceived degree of relevance, and let high-relevance information dominate low-relevance information. Information is deemed more relevant when it relates more specifically to a judged target case. Specificity is achieved either by providing information on a smaller set than the overall population, of which the target case is a member, or when information can be coded, via causality, as information about the specific members of a given population. The base-rate fallacy is thus the result of pitting what seem to be merely coincidental, therefore low-relevance, base rates against more specific, or causal, information. A series of probabilistic inference problems is presented in which relevance was manipulated with the means described above, and the empirical results confirm the above account. In particular, base rates will be combined with other information when the two kinds of information are perceived as being equally relevant to the judged case.

Consider the following problem:

Problem 1: Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are Blue, and the remaining 15% 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

* I would like to thank Amos Tversky, teacher and friend, for inspiring this study, and Reid Hastie, Baruch Fischhoff, and Sarah Lichtenstein for many constructive comments. Partial support for this research was provided by the Advanced Research Projects Agency of the Department of Defense, and was monitored by the Office of Naval Research under Contract N00014-76-C-0074 (ARPA Order No. 3052) under Subcontract 76-030-0714 from Decisions and Designs, Inc. to Perceptronics, Inc.

> and Green cabs under nighttime visibility conditions. It found that the witness was able to identify each color correctly about 80% of the time, but confused it with the other color about 20% of the time. What do you think are the chances that the errant cab was indeed Green, as the witness claimed? (Kahneman and Tversky 1972).

This is a paradigmatic Bayesian inference problem. It contains two kinds of information. One is in the form of background data on the color distribution of cabs in the city, called *base-rate* information. The second, rendered by the witness, relates specifically to the cab in question, and is here called *indicant* or *diagnostic* information.

The proper, normative way to combine the inferential impacts of base-rate evidence and diagnostic evidence is given by Bayes' rule. In odds form, this rule can be written as $\Omega = Q \cdot R$, where Ω denotes the posterior odds in favor of a particular inference, Q denotes the prior odds in favor of that particular inference, and R denotes the likelihood ratio for that inference. In the cab example above, we are interested in the probability, after the witness' testimony, that the errant cab was Green. Denote Green cabs and Blue cabs by G and B, respectively, and denote the testimony that the cab was green by g. Spelling out Bayes' Theorem in full, we obtain:

$$\Omega = \frac{P(G/g)}{P(B/g)} = \frac{P(g/G)}{P(g/B)} \times \frac{P(G)}{P(B)} = \frac{0.8}{0.2} \times \frac{0.15}{0.85} = \frac{12}{17}$$

and thus $P(G/g) = \frac{12}{12+17} = 0.41$. Note that the prior odds are based on the population base rates, whereas the likelihood ratio is determined by the indicator.

If the posterior probability of 41% seems counterintuitive to you and your initial inclination is to be 80% sure that the witness' testimony of Green is in fact reliable, then you are exhibiting the base-rate fallacy – the fallacy of allowing indicators to dominate base rates in your probability assessments. You are, however, in good company. The base-rate fallacy has been found in several experimental studies, and it manifests itself in a multitude of real-world situations.

In a 1955 paper, Meehl and Rosen warned against the insensitivity, on the part of both the designers and users of diagnostic tests, to base-

rate considerations. They lamented psychologists' proneness to evaluate tests by their hit rate (*i.e.*, diagnosticity) alone, rather than by the more appropriate measure of efficiency, which would take into account base rates, as well as costs, goals, and other relevant considerations. Clinicians are apparently unaware that they should feel less confident when a test returns a rare verdict (such as 'suicidal') than when it returns a more common one.

Such warnings persist to our day. Lykken (1975) laments current injudicious use of polygraph outputs by commercial companies, while demonstrating that even a highly accurate polygraph reading is likely to yield erroneous diagnoses when, say, it is administered to a whole population of employees, only a fraction of whom are really guilty of some offense. Dershowitz (1971), Stone (1975) and McGargee (1976) point out that since violence is a rare form of behavior in the population, base-rate considerations alone make it more likely than not that an individual who is preventively detained because he is judged to be potentially dangerous is really quite harmless, a purely statistical argument whose significance has only recently gained appreciation among jurists. Finally, Eddy (1978) has evidence of the base-rate fallacy both in the judgments of practicing physicians and in the recommendations of some medical texts.

Base rates play a problematic role in yet another legal context, namely, the fact-finding process. Though there is no definitive ruling on the status of base-rate evidence, courts are typically reluctant to allow its presentation, often ruling it inadmissible on grounds of irrelevancy to the debated issues. While some of the legal objections reflect sound reasoning, others are clearly manifestations of the base-rate fallacy. (For a discussion of base rates in the courts, see Tribe 1971.)

The counterpart of disregarding the probative impact of base rates lies in overjudging the probative impact of indicators. To hark to a well-known children's riddle, white sheep eat more grass than black sheep simply because there are more of them. Color is really no indicator of appetite – the phenomenon is a base-rate one. Similarly, the fact that in 1957 in Rhode Island more pedestrians were killed when crossing an intersection with the signal than against it (Huff 1959) does not necessarily imply that it is more dangerous to comply with traffic lights than to violate them. The indicators in these two cases seem to shoulder the diagnostic burden only because the base rates do not seem to. An entire methodology of experimental control has been conceived to guard

against this prevalent side effect of the base-rate fallacy.

The base-rate fallacy may underlie some phenomena noted in the domain of interpersonal perception as well. Nisbett and Borgida (1975) have used this notion to explain the perplexingly minimal role that consensus information typically plays in people's causal attributions, consensus data being, in effect, base-rate data. The consequences of the base-rate fallacy to interpersonal perception was also unwittingly demonstrated by Gage (1952). Gage found that predicting the questionnaire behavior of strangers drawn from a familiar population deteriorated following an opportunity to observe these strangers engaging in expressive behavior. If we suppose that (a) the indicators gleaned from these observations suppressed the base-rate information which was previously available through the familiarity with the source population of these strangers and (b) these base-rate considerations were more diagnostic (*i.e.*, more extreme) in themselves than the expressive behavior was, then Gage's results are readily understood.

Experimental studies of the base-rate fallacy

Although the existence of the base-rate fallacy has been acknowledged for quite some time (Meehl and Rosen 1955; Huff 1959; Good 1968), it was first studied in a controlled laboratory situation by Kahneman and Tversky (1973a). These investigators presented subjects with a series of short personality sketches of people randomly drawn from a population with known composition. On the basis of these sketches, subjects were to predict to which of the population subclasses the described persons were more likely to belong. Subjects were responsive to the diagnosticity of the descriptions, but they practically disregarded the fact that the different subclasses of the population were of grossly different size. Therefore, subjects were as confident when predicting membership in a small subclass (which correspondingly enjoys a smaller prior probability) as in a larger one. Kahneman and Tversky interpreted their results as showing that:

> ... people predict by representativeness, that is, they select ... outcomes by the degree to which (they) represent the essential features of the evidence ... However, ... because there are factors (*e.g.*, the prior probability of outcomes ...) which affect the likelihood of outcomes but not their representativeness, ... intuitive predictions violate the statistical rules of prediction (1973a: 237–238).

While the manner in which people derive judgments of diagnosticity from personality sketches may well proceed via judgments of representativeness, the base-rate fallacy appears in other contexts, where the representativeness argument does not seem to apply. Such, for example, is Problem 1, in which both the base rate and the indicant information is presented in numerical form. Thus, "regardless of whether or not probability is judged by representativeness, base rate information [may] be dominated" (Tversky and Kahneman 1980).

Another interpretation of Kahneman and Tversky's results was offered by Nisbett *et al.* (1976), who suggested that base-rate information is ignored in favor of individuating information, since the former is "remote, pallid and abstract", whereas the latter is "vivid, salient, and concrete" (1976: 24). However, Problem 1 presents both items of information in highly similar style – instead of being a case description, the indicant information is also statistical in nature. Thus, the phenomenon seems more general than Nisbett *et al.* may have realized.

Recent investigations have addressed themselves to the stability of the base-rate phenomenon (Lyon and Slovic 1976; Bar-Hillel 1975). A wide range of variations of the basic problem was presented to a total of about 350 subjects, including (a) changing the order of data presentation with the indicator data preceding, rather than following, the base-rate information; (b) using green rather than blue as majority color; (c) having subjects assess the probability that the witness erred, rather than the probability of correct identification; (d) having the witness identify the errant cab as belonging to the larger, rather than the smaller, of the two companies; (e) varying the base rate (to 60% and 50%); (f) varying the witness' credibility (to 60% and 50% hits); and (g) stating the problem in a brief verbal description without explicit statistics (*e.g.*, "most of the cabs in the city are Blue", and "the witness was sometimes, but rarely, mistaken in his identifications," Kahneman and Tversky 1973b). Through all these variations, the median and modal responses were consistently based on the indicator alone, demonstrating the robustness of the base-rate fallacy.

Why are base rates ignored?

The genuineness, the robustness, and the generality of the base-rate fallacy are matters of established fact. What needs now be asked is

why the phenomenon exists, *i.e.*, what cognitive mechanism leads people to ignore base-rate information in problems of Bayesian inference? The cab problem seems to rule out some possible explanations as too narrow (Nisbett and Borgida's saliency explanation), or insufficient (Kahneman and Tversky's explanation by representativeness). It is also erroneous to believe that people's failure to integrate base-rate information into their judgments reflects either lack of appreciation for the diagnostic impact of such data, or some inherent inability to integrate uncertainties from two different sources. Neither possibility is true. People demonstrate that they do appreciate the implications of base-rate information when it is the only information they have. Thus, subjects who receive a cab problem with no witness overwhelmingly chose 15% as their estimate of the probability that the hit-and-run cab was Green (Lyon and Slovic 1976; Bar-Hillel 1975). People's ability to integrate two sources of information will become apparent in their responses to some of the problems to be presented below.

The most comprehensive attempt yet to account for the base-rate fallacy is to be found in Ajzen (1977) and in Tversky and Kahneman (1980). The latter claim that "base-rate data that are given a causal interpretation affect judgments, while base-rates that do not fit into a causal schema are dominated by causally relevant data" (Tversky and Kahneman 1980: 50). They then proceed to demonstrate this claim using a number of inference problems, some based on Bar-Hillel (1975).

The causality argument, I believe, is incomplete. (a) It is again too narrow. I shall show that under some conditions, even non-causal base-rate information will affect judgments. Thus, although "people rely on information perceived to have a causal relation to the criterion", it is not always the case that they "disregard valid but non-causal information" (Ajzen 1977: 303). (b) The causality argument accounts for *when* rather than *why* base rates will be ignored. In this paper, I shall attempt to give a dynamic account for the base-rate fallacy. Causality will be but one factor in this more general account.

The central notion in the proposed account is the notion of *relevance*. I believe that subjects ignore base-rate information, when they do, because they feel that it *should* be ignored – put plainly, because the base rates seem to them *irrelevant* to the judgment that they are making. This notion, which will be presently elaborated and supported by data, was initially based on some introspection and some anecdotal evidence.

Subjects, when occasionally queried informally about their erroneous response to the cab problem, vehemently defended their witness-based judgment, denying that the cab distribution should have "had anything to do with" their answer. Lyon and Slovic (1976) presented subjects with a forced-choice question regarding the relevance of the two items of information in the cab problem. Subjects were offered reasoned statements in favor of (a) only base rates being relevant; (b) only the indicator being relevant; and (c) both being relevant. In spite of the fact that the correct argument was explicitly formulated in (c), 50% of their subjects chose (b). In another study, Hammerton (1973) gave his subjects a similar kind of problem, but omitted the base rates altogether. His subjects showed no awareness that a vital ingredient was missing.

I propose that people may be ordering information by its perceived degree of relevance to the problem they are judging. If two items seem equally relevant, they will both play a role in determining the final estimate. But if one is seen as more relevant than the other, the former may dominate the latter in people's judgments. Since less relevant items are discarded prior to any considerations of diagnosticity, an item of no diagnostic value, if judged more relevant, may dominate an item of high diagnosticity. (This is similar to the way 'relevance' is used in a court of law. Evidence considered 'irrelevant' by the judge is not admitted – though clearly the side wishing to introduce it thinks it will have an impact on the judge or jury – whereas often 'relevant' evidence without any diagnostic impact is admitted.) These levels of relevance are crude, almost qualitative, categories. And it is only *within* levels that judged diagnosticity will affect the weights assigned to different pieces of information.

A crucial question is, of course, what determines the ordering of items by relevance. While I cannot offer a comprehensive answer at this point, it seems to me that one factor affecting relevance is specificity. Thus, if you have some information regarding some population, and other information regarding some subset of that population, then the latter is more relevant than the former for making judgments about a member of that subset. Often such more specific information should normatively supercede the more general information. Clearly, for example, the base rate of married people among young female adults should be used in place of the base rate of married people in the entire adult population when judging the marital status of a young female adult.

In other cases, some examples of which are the Bayesian inference problems in which the base rate fallacy is manifest, this is not so. To see why not, consider for example the cab problem. It contains some general base rate information (*e.g.*, "85% of the cabs in the city are Blue") and some indicant information pertaining more specifically to the judged case. The indicator in the cab problem, *i.e.*, the witness' testimony, actually focuses directly on the unique cab involved in the accident, and identifies it as green. It is, however, not perfectly reliable. This, then, is one example where more specific information, though it seems more relevant, should not supercede the base rate. Another example will be encountered in Problem 3 below. There, the more specific information does not concern the same predicate as the more general one, though it may seem to. Thus, it not only shouldn't supercede the general information, but it should be completely ignored.

Specificity can be brought about in several ways. The most straightforward one is to give information about some subset of the population. When this subset contains only the judged case, it has been called "individuating" information (Kahneman and Tversky 1973a). Information which testifies directly (if imperfectly) about some member of the population, such as the witness in the cab problem or Lyon and Slovic's mechanical device, will be called *identifying information*. An indirect way of achieving specificity is via causality. Causality provides a means of attaining specific, individual characteristics from population characteristics. If one is told, for example, that "85% of cab accidents in the city involve [blue] cabs" (Tversky and Kahneman 1980), this population base rate is readily interpretable as saying that blue cabs are more accident-prone than green cabs – an interpretation which makes the base-rate more specifically related to the accident likelihood of individual cabs. Thus, the causality factor identified by Tversky and Kahneman (1980) is just one means of enhancing relevance.[1] And it is relevance, rather than causality *per se*, which determines whether or not base rates will be incorporated into probability judgments. More precisely, it is relative relevance – *i.e.*, two items of information will be

[1] A graphic example may be in order here: imagine a grid of squares. There are many ways of coloring it which will turn 80% of the grid's area red and 20% green. One is to color 80% of the squares all red, and 20% all green. Another is to color each square 80% red and 20% green. Yet another is to color some of the squares using one mixture, and others using another mixture. In the second case, but not in the first, the grid color statistics apply to each square. Causality, I claim, is a way of inferring 'square' characteristics from 'grid' characteristics.

integrated if they are perceived as equally relevant. Otherwise, the more relevant one will dominate the less relevant one.

This study will unfold as follows: first, I shall present a number of problem prototypes in which the base rate fallacy is manifested: (1) the paradigmatic Cab Problem, in which a *coincidental* base rate (*i.e.*, one which cannot be causally interpreted) is pitted against identifying, but not perfectly reliable, information; (2) the Suicide Problem, in which a coincidental base rate is pitted against information which, though it is a base rate, can be causally linked to the judged outcome; (3) the Dream Problem, which gives two base rates in a non-Bayesian inference problem; and (4) the Urn and Beads Problem, in which a more general base rate is pitted against seemingly more specific base-rate information. I shall then show that when base rates are not judged less relevant than indicant information, they are not ignored. In the Intercom Problem, coincidental base rates are pitted against coincidental indicant information. In the Motor Problem, causally-interpreted base rates are pitted against identifying, but not perfectly reliable, information. According to the proposed account, the base rates should not be judged less relevant than the indicant information in both these problems, and therefore the base rate fallacy should disappear.

The Experiments

Subjects and method

The empirical core of this paper is a collection of inference problems, like Problem 1, which were presented to a total of about 1500 *S*s. Except for a small number of undergraduate volunteers, the *S*s were predominantly Hebrew University applicants who answered the questions in the context of their university entrance exams, and thus presumably were highly motivated to do their best. *S*s usually received only one problem, but occasionally two or three. When *S*s received more than one problem, these were chosen to be quite different from each other, so as to minimize interference. The total number of responses analyzed approaches 3000. The *S*s were all high school graduates, mostly 18–25 years old, and of both sexes. The *S*s were not instructed to work quickly, but questionnaires were retrieved after about four minutes (per question), and those who had not answered by then were simply discarded. This was ample time for almost all of the *S*s.

In all, about 45 problems were employed (see Bar-Hillel 1975), only a small subset of which will be presented in detail. The rest will be only briefly sketched.

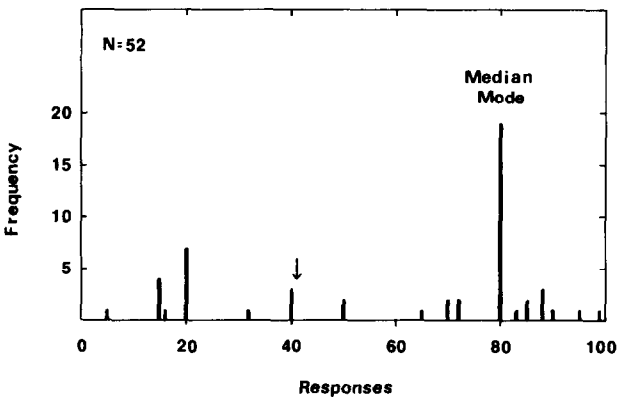

Fig. 1. Distribution of responses to the Cab Problem. In this figure, as in those to follow, the arrow indicates the correct Bayesian estimate; the median and modal responses are also shown.

The Cab Problem

Problem 1, with which this paper opened, serves as a point of departure for much of the discussion of the base-rate fallacy. Note that it offers a coincidental base rate and not–perfectly–reliable identifying information.

Fig. 1 presents the distribution of estimates that 52 *S*s gave to this problem.[2] Thirty-six percent of these *S*s based their estimate on the witness' credibility alone (80%), ignoring the base rate altogether. Eighty percent was also the median estimate. Only about 10% of the *S*s gave estimates that even roughly approximated the normative Bayesian estimate of 41%.

The same pattern of results was obtained with the whole spectrum of variations described in the introductory section. The modal answer, which invariably matched the witness' diagnosticity, was given by up to 70% of the *S*s.

The Cab Problem results, taken alone, would not necessarily have justified the name 'base-rate fallacy', since another error, unrelated to that of overlooking pertinent information, could account for them. Suppose, for instance, that in spite of the careful wording of the problem, *S*s confuse the information that "the witness (was) able to identify each color correctly about 80% of the time", formally coded as $P(g/G) = P(b/B) = 80\%$, with "80% of each of the witness' color identifications turn out to be correct", formally coded as $P(G/g) = P(B/b) = 80\%$. Such an interpretation, to be sure, is unwarranted, not merely by the formulation of the problem, but also because a very bizarre perceptual mechanism would have to be assumed to produce $P(G/g) = P(B/b) = 80\%$ in arbitrary base-rate conditions, given that we take

[2] Ninety-five additional subjects were given this problem by Kahneman and Tversky (1972) and by Lyon and Slovic (1976), with similar results.

Problem 2: A study was done on causes of suicide among young adults (aged 25 to 35). It was found that the percentage of suicides is three times larger among single people than among married people. In this age group, 80% are married and 20% are single. Of 100 cases of suicide among people aged 25 to 35, how many would you estimate were single?

The distribution of estimates that 37 *S*s gave to Problem 2 is shown in fig. 2. Forty-three percent of the *S*s gave a response (75%) based on the indicant information alone (*i.e.*, 3:1), completely ignoring the fact that more young adults are married than single. The median response was also 75%.

A Bayesian estimate based on the given data gives the answer as 43% (*i.e.*, $\Omega = \frac{0.2}{0.8} \times 3 = \frac{3}{4}$), but only six responses fell between 30% and 50%.

To test for robustness, Problem 2 was subjected to a host of variations. These included (a) not mentioning the base rates explicitly within the problem (presumably all our *S*s knew that a majority of adults aged 25 to 35 are married); (b) asking *S*s to supply, along with their answers, estimates of the missing, but necessary, base rate (the results of these estimates confirmed the assumption in [a][4]); (c) varying the base rates (using the values 50%, 10%, and 5%) – this was done, respectively, by partitioning the population into males *vs.* females, only children *vs.* siblings, and people with a known history of depression *vs.* 'normal' people; (d) varying the likelihood ratio (to 9); (e) providing the purported suicide rates themselves (5% and 15% of deaths) rather than just their likelihood ratios; (f) inverting the indicator to support rather than contradict the base-rate implication; (g) asking about the chances than an individual suicide was single, rather than for the number of singles among 100 suicides; and (h) changing the contents of the cover story from suicide rates to dropout rates among male and female students in the Hebrew University Medical School.

The 14 problems produced by these variations did not form a factorial design, as different problems incorporated different numbers of the listed variations. In all, they were presented to some 600 *S*s. The modal response was 75% throughout.[5] It was given by between 25% and 80% of the respondents. The median response was 75% in ten of the problems and 70% in three. The Suicide Problem, therefore, replicated the results obtained in the Cab Problem, without being subject to the same criticism.

Actually, if the Suicide Problem results were taken alone, they too could be explained without resort to the base-rate fallacy. Just read "the *number* of suicides

[4] According to the Israel Bureau of Statistics, 85% of the 25–35 age group in Israel (where this study was run) are married. However, since subjects estimate this proportion as 80% (median and modal response of 32 *S*s, with an interquartile range of 70–80%), I used a proportion conforming to their guess.

[5] Except in the one problem where the likelihood ratio was 9. There the median response was 90%, and the mode 80%.

percepts to be caused by external events and not *vice versa*.[3] Nevertheless, *S*s may confuse the two, and if so, their error is not that of overlooking pertinent information. If you believe you are told that $P(G/g) = 80\%$, *i.e.*, that when the witness says "the cab was Green" (or Blue, for that matter), he stands an 80% chance of being correct, then you are quite right in giving 80% as your answer, irrespective of what the base-rate conditions are. Many of the contexts in which the base-rate fallacy has been manifested are candidates for the same criticism, since they resemble the Cab Problem in offering indicant information which seems to be actually making your judgment for you, albeit with less than perfect reliability.

We need not concern ourselves overly with this point, however, since the base-rate fallacy is readily evident in a different type of problem, where no identifying information exists. Such is the following problem.

The Suicide Problem

Formally, the following problem resembles Problem 1. It is a Bayesian inference problem, with a prior derived from a base rate. However, the indicant information here is also presented actuarially, in the form of a likelihood ratio of some property in two population subclasses.

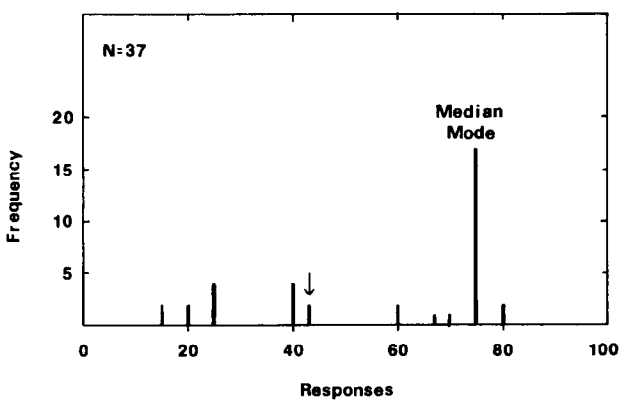

Fig. 2. Distribution of responses to the Suicide Problem, Problem 2.

[3] Only under conditions of uniform base rates does the claim that each color has an equal chance of being identified properly entail the claim that each color attribution has an equal chance of turning out to be correct. It is for this reason, of course, that the diagnosticity of indicators is typically stated in terms of their Hit and Correct-reject rates, rather than in terms of their efficiency, as Meehl and Rosen (1955) would have it. It is the former, but not the latter, which, being a constant feature of the indicator, remains invariant under fluctuating base rates, costs, *etc.*

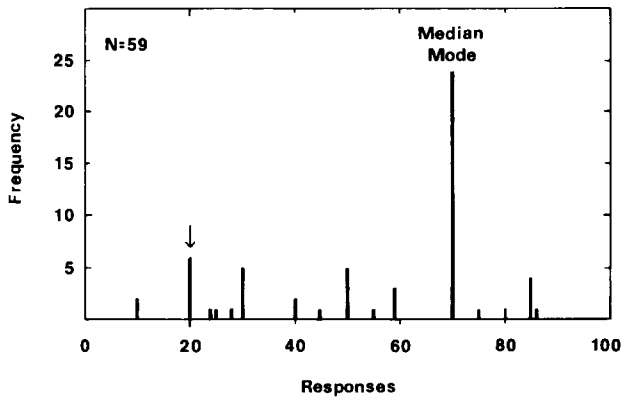

Fig. 3. Distribution of responses to the Dream Problem, Problem 3.

base rate of dreaming alone. Psychologically speaking, the data seem to tell the converse story. Never mind the individual base rate for dreaming in the overall population. Mrs. X is married, hence we should look at the statistics of married couples, which tell us that when people marry, they tend to find similarly classified mates. For a married target case, therefore, the base rate for matching among couples should predominate. That the results support this analysis can be seen in fig. 3.

Why do the matching statistics loom as more relevant to the judgment than the dreaming statistics? Causality does not seem to be the reason, since one can as readily derive from the dreaming statistics the implication that people tend to dream, as from the matching statistics the implication that they tend to match their spouses. But married people form a subset of the entire adult population. Therefore, statistics about that group are more specifically related to Mr. X than the overall population statistics. It is this specificity that enhances the relevance of the matching statistics over the dreaming statistics.

To demonstrate more firmly that the matching statistics prevail for reasons of specificity and not of causality, consider Problem 3′:

> *Problem 3′*: Studies of dreaming have shown that 80% of people of both sexes report that they dream, if only occassionally, whereas 20% claim they do not remember ever dreaming. Accordingly, people are classified by dream investigators as 'Dreamers' or 'Nondreamers'. With respect to dreaming, mating is completely random.
>
> Mrs. X is a Nondreamer.
>
> What do you think are the chances that her husband is also a Nondreamer?

This formulation clearly removes any causal link between the classification of husband and wife. Some other formulations used in variations on Problem 3′ were: "The classification of husband and wife was found to be independent" and "the

is three times larger among single people than among married people" for "the *percentage* . . .". Note, however, that the same response pattern was obtained when the suicide percentages were stated explicitly. In general, 'carelessness' explanations of the base-rate fallacy should not be pushed too far unless the same, or highly similar, confusions can account for all the results. Finding an *ad hoc* reformulation for each type of problem is too much like finding a question to fit the answer. The base-rate fallacy, I believe, is not a side effect of some other error, but an error unto itself.

How can the Suicide Problem results be explained using the relevance notion? What makes one base rate dominate another? The fact that some property (*e.g.*, suicide rate) is distributed differently in two population subclasses (*e.g.*, single *vs.* married) is a powerful invitation, psychologically, for a causal interpretation. The differential suicide rates readily suggest that marital status is causally linked to suicide; via loneliness, perhaps, if the rate is given as higher for singles; via the frustrations of married life, perhaps, if it is given as higher for marrieds. Once a causal link is established between one item of information and the judged outcome, its relevance is enhanced, and it overrides the base rate seen as merely statistical (*vis-à-vis* the judged outcome).

The Dream Problem

Problem 2 generalized the base rate fallacy from identifying information to actuarial indicant information. The following problem generalizes it outside a Bayesian framework altogether.

> *Problem 3*: Studies of dreaming have shown that 80% of adults of both sexes report that they dream, if only occassionally, whereas 20% claim they do not remember ever dreaming. Accordingly, people are classified by dream investigators as 'Dreamers' or 'Nondreamers'. In close to 70% of all married couples, husband and wife share the same classification, *i.e.*, both are Dreamers or both are Nondreamers, whereas slightly more than 30% of couples are made up of one Dreamer and one Nondreamer.
>
> Mrs. X is a Nondreamer.
>
> What do you think are the chances that her husband is also a Nondreamer?

In this problem two base rates are offered, that of dreaming for individuals, and that of matching for married couples. The target case is a married individual, so both base rates apply to him. Ostensibly, the two items play analogous roles. Undoubtedly, if either were given alone, it would have determined the majority of responses. In fact, however, there is a marked asymmetry between the two items, from both a formal and a psychological point of view. Formally speaking, only the rate of dreaming is relevant to the judgment requested, since the data tell us that mating is random. We expect 64% of couples (0.80 X 0.80) to be both Dreamers, and 4% (0.20 X 0.20) to be both Nondreamers, for a total of 68% (*i.e.*, 'close to 70%'). Either of the base rates given is equivalent to random mating, given the other base rate. Thus, a spouse's classification is entirely irrelevant – assessments should be based on the

The appearance of identical modal estimates in the first two rows and in the last two rows reflects insensitivity to priors, *i.e.*, the base-rate fallacy, since answers vary with the sample composition, but not with the urn population. These problems show once more that relevance can be mediated by more than just causality. Here the sample information overrides the urn-population information. It seems strained, at least, to say that the sample composition is causally related to the distribution of beads in the urns, but not to the distribution of the urns themselves. It is, after all, the result of a two-step sampling procedure. A case can be made, however, for the greater relevance of within-urn composition over between-urn composition via specificity. We (the *S*s) already *know* what the sample looks like, so the procedure that generated it is irrelevant. What needs to be done is to look at the sample composition and see what it tells us about its origin (*i.e.*, how well it represents the different possible urns).

In the four problems presented above I have argued that people attended to one item of information and disregarded the other because the latter seemed less relevant. To test this directly, these four problems were given to *S*s who were asked which of the two items presented in each problem they "think is more relevant for guessing" the target case class membership: the base rate (*e.g.*, "the proportion of Blue *vs.* Green cabs"), the indicant information (*e.g.*, "the suicide rate among single *vs.* among married people"), or both equally. Sixty-eight percent of 94 *S*s picked the indicator as more relevant, and only 12% judged both items equally relevant.

A revised Cab Problem

To conclude this set of problems in which the base rate fallacy is manifested, I now present a modified version of the original Cab Problem. Here the witness is replaced by actuarial, but more specific, information. As we expect, it dominates the more general base rate. The novelty here is in the fact that this is the first problem presented in which the base rate 'fallacy' is no fallacy: it is quite appropriate to disregard the first base rate.

Problem 5: Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green.

A cab was involved in a hit-and-run accident at night.

The police investigation discovered that in the neighborhood in which the accident occurred, which is nearer to the Green Cab company headquarters than to the Blue Cab company, 80% of all taxis are Green, and 20% are Blue.

What, do you think, are the chances that the errant cab was green?

Of 37 *S*s who received this problem, almost 60% gave an estimate of 80%. The overall pattern of results resembles that obtained in the different variations of the previous problems. This may give us some insight into what people are doing in those problems. The relevance notion accounts only for the responses of the median and modal *S*, so some of the variance in the *S*s' responses remains unexplained. The

spouse's classification was found to have no predictive validity". In some cases, the cover story was also changed, to couples of mother-daughter (rather than husband-wife). A total of 270 *S*s saw some version of Problem 3′. The median response was always 50%. The modal response was 50% in five questions, and 20% in the two others. Note that 50% would have been a reasonable answer if no base-rate were given. But in the presence of the base-rate it means that the *S*s were basing their answers on totally worthless information.

Urn and Beads Problem

The next family of problems was modeled after Edwards (1968). For many years, researchers working within the Bayesian approach to information integration were content to conclude that "the subjects' revision rule is essentially Bayes' Theorem" (Beach 1966: 6; see also Edwards 1968; Peterson and Beach 1967; Schum and Martin 1968). The advent of Kahneman and Tversky's judgmental-heuristics approach spelled the demise of the Bayesian approach. The following urn-and-beads problem again shows that people here are not poor Bayesians, but rather non-Bayesians.

> *Problem 4*: Imagine ten urns full of red and blue beads. Eight of these urns contain a majority of blue beads, and will be referred to hereafter as the Blue urns. The other two urns contain a majority of red beads, and will be referred to hereafter as the Red urns. The proportion of the majority color in each urn is 75%. Suppose someone first selects an urn on a random basis, and then blindly draws four beads from the urn. Three of the beads turn out to be blue, and one red.
>
> What do you think is the probability that the beads were drawn from a Blue urn?

In other versions of this question, the number of Blue urns was given as five out of the total ten, and/or the number of blue beads in the sample was given as one. Results are presented in table 1.

Table 1
Summary of Urn and Beads Problem.

	Problem description				Results			
Number	Urns		Sample		Normative Bayesian assessment	Modal assessment	Freq. of modal assessment	No. of subjects
1	8B	2R	3B	1R	0.97B	0.75B	14	54
2	5B	5R	3B	1R	0.90B	0.75B	20	50
3	8B	2R	1B	3R	0.31B	0.25B	13	53
4	5B	5R	1B	3R	0.10B	0.25B	6	20

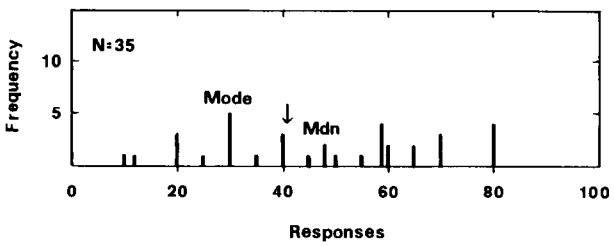

Fig. 4. Distribution of responses to the Intercom Problem, Problem 7.

by lowering the relevance of the indicant information. Only with great contrivance can this information be viewed as more causal than the base rate, and it certainly isn't more specific.

Problem 7: Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are Blue, and 15% are Green. A cab was involved in a hit-and-run accident at night, in which a pedestrian was run down. The wounded pedestrian later testified that though he did not see the color of the cab due to the bad visibility conditions that night, he remembers hearing the sound of an intercom coming through the cab window. The police investigation discovered that intercoms are installed in 80% of the Green cabs, and in 20% of the Blue cabs.

What do you think are the chances that the errant cab was Green?

Fig. 4 shows the distribution of 35 *Ss*' responses to this problem. The attempt to lower the relevance of the indicant information was apparently successful, for the indicant information did not dominate the base rates. The median response is 48%, close to the correct 41%. These results differ from earlier ones not only in the median response, but in the entire distribution. It is flatter, 'noisier', suggesting that there is no prevailing strategy of integration favored by a large proportion of the *Ss*. A similar pattern of results emerged in a second attempt (not reproduced here) to construct a problem along the same lines as Problem 7, *i.e.*, with indicant information which could hardly be judged more relevant to the outcome than was the overall base rate.[6] The response distribution of 23 *Ss* had a median of 42%, and no unique mode.

A second strategy for eliminating the base rate fallacy lies in enhancing the perceived relevance of base rates to make them seem as relevant as indicant information typically is. This was attempted in Problem 8, by making the base rates both causally related to the judged outcome, and more specific:

Problem 8: A large water-pumping facility is operated simultaneously by two giant motors. The motors are virtually identical (in terms of model, age, *etc.*),

[6] In this problem a bookstore was described, with 85% English books and 15% Hebrew ones. Eighty percent of Hebrew books and 20% of English ones are hard cover, the rest soft cover.

amount of unexplained variance is not, however, much greater there than in Problem 5, where all responses, from both a normative and psychological viewpoint, should have converged on 80%. I take this as further evidence that people just don't know when specific information should replace more general information, and when it should only modify it.

Integration of equally relevant items

It might seem at this point that people's failure to integrate base-rate information into their judgments reflects some inherent inability to integrate uncertainties from two different sources. According to the proposed account, however, such 'inability' should only be apparent when two items are not equally relevant for some judgment. Let us see what happens when two pieces of information *are* equally relevant.

One obvious way of making two items appear equally relevant lies in letting them play entirely symmetrical roles in a problem. Such are the roles of the two witnesses in the following problem.

> *Problem 6*: Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green.
>
> A cab was involved in a hit-and-run accident at night.
>
> There were two witnesses to the accident. One claimed that the errant cab had been Green, and the other claimed that it had been Blue. The court tested the witnesses' ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found the first witness (Green) able to identify the correct color about 80% of the time, confusing it with the other color 20% of the time; the second witness (Blue) identified each color correctly 70% of the time, and erred about 30% of the time.
>
> What do you think are the chances that the errant cab was Green, as the first witness claimed?

Of 27 *S*s responding to Problem 6, 14 gave an assessment of 55% (midway between the assessments implied by each witness alone, disregarding base rates), and all but one gave assessments between 50% and 60%.

In another problem (not reproduced here) *both* witnesses identified the cab as Green. Twenty-four of the 29 *S*s answering that problem gave an assessment of 75% – again, midway between the two witness-based assessments. While these *S*s were still disregarding the base rate, they appear to have been averaging the probabilistic implications of the two testimonies. Averaging probabilities is, of course, not the proper way to calculate the joint impact of the two independent testimonies. But it does clearly indicate that both sources are being considered.

If making two items of information appear equally relevant ensures that they will both be somehow integrated judgmentally, then all one needs to do to overcome the base-rate fallacy in a Bayesian inference problem is to find a way of equating the relevance of base rates and indicant information. Problem 7 does this

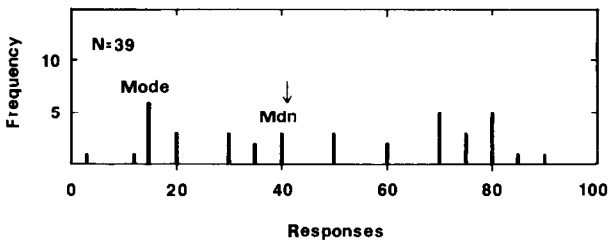

Fig. 5. Distribution of responses to the Motor Problem, Problem 8.

was more relevant for guessing the target case's class membership. Almost 40% of 50 *S*s indicated that they thought both were equally likely, and the rest were just about equally divided between base rates and indicators. Contrast this with the 12% who thought both items were equally likely in Problems 1 to 4, and 68% who marked the indicators as more relevant ($\chi^2_{2df} = 15.74$, $p < 0.001$).

Discussion

This study presented evidence for an explanation of the base-rate fallacy based on a notion of relevance: people integrate two items of information only if both seem to them equally relevant. Otherwise, high relevance information renders low relevance information irrelevant. One item of information is more relevant to a judged case than another if it somehow pertains to it more specifically. Two means whereby this can be achieved were pointed out: (1) The dominating information may refer to a set smaller than the overall population to which the dominated item refers, but of which the judged case is nevertheless a member. (2) The dominating information may be causally linked to the judged outcome, in the absence of such a link on behalf of the dominated information. This enhances relevance because it is an indirect way of making general information relate more specifically to individual cases. This proposal goes beyond the "causal schema" offered by Tversky and Kahneman (1980) by showing causality to be a special case of a more general notion, that of relevance.

My approach to enhancing relevance has been to manipulate the contents of problems subjects face in ways which I believe affect their judgment of relevance, which I shall call *internal* relevance. An alternative way is to make people aware of the relevance of some item *externally*,

except that a long history of breakdowns in the facility has shown that one motor, call it A, was responsible for 85% of the breakdowns, whereas the other, B, caused 15% of the breakdowns only.

To mend a motor, it must be idled and taken apart, an expensive and drawn out affair. Therefore, several tests are usually done to get some prior notion of which motor to tackle. One of these tests employs a mechanical device which operates, roughly, by pointing at the motor whose magnetic field is weaker. In 4 cases out of 5, a faulty motor creates a weaker field, but in 1 case out of 5, this effect may be accidentally caused.

Suppose a breakdown has just occurred. The device is pointed at motor B.

What do you think are the chances that motor B is responsible for this breakdown?

As in the Cab Problem 1 and other instances of imperfect diagnosis, we have here a device that singles out a specific motor as the likely cause of a mechanical failure, *i.e.*, identifying information. However, the present base rate is readily interpreted as an individual attribute of the two motors, implying that one motor, A, is in worse shape than the other. Thus, both the base rate and the indicator single out, albeit probabilistically, a specific suspect.

Apparently *S*s indeed took both items to be equally relevant, since the pattern of results given by 39 *S*s to this question is similar to that obtained in Problem 7. Over 60% of the *S*s gave assessments interpretable as weighted averages of the two items of information (*i.e.*, they lie strictly between 15% and 80%, the assessments corresponding to the individual items), and the median of the distribution is at 40%, remarkably close to the correct Bayesian posterior of 41%.

Problem 8 was one of five problems in which base rates were made more relevant. They all used the same parameters and format of presentation, and differed in cover story only.[7] They, as well as Problem 7 and its one variant, were all characterized by the following results: the median always lay strictly between 15% and 80% (ranging from 38% to 75%, with three of the medians within 3 percentage points of 41%, the normative answer). The mode, if there was a unique one, was shared in one case by 40% of the *S*s (but it was 15%, not 80%), and otherwise was shared by less than 30% of the *S*s.

A final piece of evidence that shows that Problems 7 and 8 were successful in making base rates and indicators appear equally relevant lies in the responses given by *S*s who were asked to indicate which of the two items of information in them

[7] I have not reproduced them for space considerations, but they are available upon request. In some of them it seems reasonable to assert that the base rates were specific, but not causal. In Kahneman and Tversky's Problem 10 (1980), the base rate was causal but not specific. In the present Problem 8 it was probably both. In all of them, the indicant information was of the identifying kind. Kahneman and Tversky (1980) also report the results of a variant of the Suicide Problem where both items of information are base rates, and both are causally linked to the outcome: the overall base rate is the proportion of male *vs.* female adolescents among attempted suicides, and the indicant information is the differential rates of successful attempts in the two sexes (Problem 12).

for example, by letting the same subject make judgments on a series of problems which differ only in the value of some item of information. Such a strategy enhances the salience of this information, and thus possibly makes it more externally relevant. This approach was utilized with modest success by Fischhoff *et al.* (1979); I view their results as additional support for the notion of relevance.

The empirical core of the study consisted of a series of several problem-types, each representing a larger family of problems that were used in the course of the study. The problems (a) established the robustness of the base-rate fallacy, by replicating results over many variations; (b) provided counter examples to the accuracy or generality of some possible accounts of the fallacy; and (c) directly confirmed some implications derived from the account herewith put forth, most significantly that base rates *can* be made to influence subjective probability judgments.

The problems studied can be roughly divided into two groups. The first group contains problem types in which one item was more relevant, and therefore dominated another (1, 2, 3, 3′, 4, 5, and their variants). These problems are characterized by a relatively high degree of consensus among subjects, with responses converging on the response implied by the more relevant item. The problems in the second group, on the other hand (6, 7, 8), yielded flatter, less elegant distributions, with two or more modes (Problem 6 is an exception). They are more aptly described as having no apparent dominance rather than as problems in which a well-defined integration policy emerged. In this latter group, both items were designed to appear equally relevant to subjects, and the base-rate fallacy was no longer in evidence.

Several questions are raised by the problems and their results: what happens when a general but causal base rate is pitted against more specific but non-causal information? is it possible to reverse the base-rate fallacy – *i.e.*, given a causal base rate and a coincidental indicator, will the base rate systematically dominate the indicant information? when people *are* being influenced by both items, how do they go about integrating? how can greater consensus be achieved here? greater validity?

There are some formal challenges that are posed by these problems as well: how *should* uncertainties from more than one source be combined? The Bayesian model offers an answer in some cases, but surely not all. Even the two witness problem cannot be dealt with by Bayes' Theorem without an assumption of conditional independence of the

witnesses. How should, for example, the rates of dreaming in two sets (say among the elderly, and among females) be combined to assess the rate of dreaming in their intersection (*i.e.*, an elderly woman)? These problems, empirical and formal, call for further research.

I believe that this study deepens our understanding of the causes of the base-rate fallacy, and describes conditions under which it will not be manifest. It is important to remember, however, that in the typical Bayesian reasoning contexts which people encounter in daily life, there is every reason to expect the fallacy to operate. Psychologists are familiar with the fact that as information is added in a probabilistic inference task, confidence increases rapidly, whereas accuracy increases only minimally, if at all (Oskamp 1965). The base-rate fallacy is a demonstration of how new information may actually lead to a *decline* in predictive performance, by suppressing existing information of possibly greater predictive validity. In the mind of the human judge, more is not always superior to less.

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., 1975. The base-rate fallacy in probability judgments. Doctoral dissertation presented to the Hebrew University, Jerusalem.

Beach, L. R., 1966. Accuracy and consistency in the revision of subjective probabilities. IEEE Transactions on Human Factors in Electronics 7, 29–37.

Dershowitz, A., 1971. Imprisonment by judicial hunch. American Bar Association Journal 57, 560–564.

Eddy, D., 1978. Personal communication.

Edwards, W., 1968. Conservatism in human information processing. In: B. Klienmuntz (ed.), Formal representation of human judgment. New York: Wiley.

Fischhoff, B., P. Slovic and S. Lichtenstein, 1979. Subjective sensitivity analysis. Organization Behavior and Human Performance 23, 339–359.

Gage, N. L., 1952. Judging interests from expressive behaviour. Psychological Monographs 66 (Whole No. 350).

Good, I. J., 1968. Statistical fallacies. In: The international encyclopedia for the social sciences, vol. 5. MacMillan and Free Press, pp. 292–301.

Hammerton, M., 1973. A case of radical probability estimation. Journal of Experimental Psychology 101, 242–254.

Huff, D., 1959. How to take a chance. Harmondsworth: Pelican Books.

Kahneman, D. and A. Tversky, 1972. On prediction and judgment. Oregon Research Institute Bulletin 12 (4).

Kahneman, D. and A. Tversky, 1973a. On the psychology of prediction. Psychological Review 80, 237–251.

Kahneman, D. and A. Tversky, 1973b. Unpublished data.
Lykken, D. T., 1975. The right way to use a lie detector. Psychology Today 8, 56–60.
Lyon, D. and P. Slovic, 1976. Dominance of accuracy information and neglect of base rates in probability estimation. Acta Psychologica 40, 287–298.
McGargee, E. I., 1976. The prediction of dangerous behavior. Criminal Justice and Behavior 3, 3–22.
Meehl, P. and A. Rosen, 1955. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin 52, 194–215.
Nisbett, R. E. and E. Borgida, 1975. Attribution and the psychology of prediction. Journal of Personality and Social Psychology 32, 932–943.
Nisbett, R. E., E. Borgida, R. Crandall and H. Reed, 1976. Popular induction: information is not necessarily informative. In: J. S. Carroll and J. W. Payne (eds.), Cognition and Social Behavior. Hillsdale, N. J.: Lawrence Erlbaum Associates.
Oskamp, S., 1965. Overconfidence in case-study judgments. Journal of Consulting Psychology 29, 261–265.
Peterson, C. R. and L. R. Beach, 1967. Man as an intuitive statistician. Psychological Bulletin 68, 29–46.
Schum, D. A. and D. W. Martin, 1968. Human processing of inconclusive evidence from multinomial probability distributions. Organizational Behavior and Human Performance 3, 353–365.
Stone, A. A., 1975. Mental health and law: a system in transition. Washington, D. C.: Government Printing Office.
Tribe, L. H., 1971. Trial by mathematics: precision and ritual in the legal process. Harvard Law Review 84, 1329–1393.
Tversky, A. and D. Kahneman, 1980. Causal schemas in judgments under uncertainty. In: M. Fishbein (ed.), Progress in social psychology, vol. I. Hillsdale, N. J.: Lawrence Erlbaum Associates.