Journal of Experimental Psychology:
Human Perception and Performance
1988, Vol. 14, No. 3, 513-525

Copyright 1988 by the American Psychological Association, Inc.
0096-1523/88/$00.75

Presentation and Content: The Use of Base Rates as a Continuous Variable

Gerd Gigerenzer, Wolfgang Hell, and Hartmut Blank
University of Constance, Constance, West Germany

Do subjects, in probability revision experiments, generally neglect base rates due to the use of a representativeness heuristic, or does the use of base rates depend on what we call the *internal problem representation*? In Experiment 1, we used Kahneman and Tversky's (1973) engineer–lawyer problem, where random sampling of descriptions is crucial to the internal representation of the problem as one in probability revision. If random sampling was performed and observed by the subjects themselves, then their judgments conformed more to Bayesian theory than to the representativeness hypothesis. If random sampling was only verbally asserted, judgments followed the representativeness heuristic. In Experiment 2 we used the soccer problem, which has the same formal structure but which the subjects' every day experience already represents as a probability revision problem. With this change in content, subjects' judgments were indistinguishable from Bayesian performance. We conclude that by manipulating presentation and content, one can elicit *either* base rate neglect *or* base rate use, as well as points in between. This result suggests that representativeness is neither an all-purpose mental strategy nor even a tendency, but rather a function of the content and the presentation of crucial information.

From its origins circa 1660 until the mid-nineteenth century, probability theory was closely identified with rational thinking. In Laplace's famous phrase, probability theory was believed to be "only common sense reduced to calculus" (Laplace, 1814/1951, p. 196). For the classical probabilists, their calculus codified the intuitions of an elite of reasonable men in the face of uncertainty. And if these reasonable intuitions deviated from the laws of probability theory, it was the latter that were cast into doubt. Such discrepancies actually influenced the way in which probability theory developed mathematically (Daston, 1980). In the early decades of the nineteenth century, probability theory shifted from being a description of the intuitions of rational individuals to one of the behavior of the irrational masses (Porter, 1986). But in the 1960s and 1970s experimental psychology reestablished the link between probability theory and rational thinking under uncertainty.

However, the new alliance differed from the old in two important respects. First, it was now probability theory, rather than intuitive judgments, that was the normative standard. Although probabilists have from time to time doubted whether the additivity law holds in all cases (Shafer, 1978), and although there is evidence that different statistical approaches suggest different answers to the same problem (Birnbaum, 1983), psychologists have generally assumed that statistics spoke with one voice—a necessary assumption for the new normative approach. Second, the link between probability theory and human thinking has become the subject of experimental research. First, by using urn-and-balls problems (e.g., Edwards, 1968; Phillips & Edwards, 1966) and then more realistic problems (e.g., Kahneman, Slovic, & Tversky, 1982), psychologists concluded from these experiments that human judgments, even those of experts (Tversky & Kahneman, 1971), deviate considerably from probabilistic rules like Bayes's theorem. Seen from the new normative perspective, such deviations seemed to indicate widespread irrationality in human thinking under uncertainty.

Here we are concerned with the empirical discrepancy rather than the implications of the normative viewpoint, and we shall discuss such discrepancies in the context of Bayesian-type problems of probability revision. We want to distinguish at the outset between base rate neglect and base rate fallacy. *Base rate neglect* is a descriptive term that carries no implication as to whether base rates *should* be used in a given case. *Base rate fallacy*, on the other hand, is a normative term that asserts not only that base rates were neglected but also that they should not have been. We shall report two experiments that try to reduce base rate neglect to null, in order to learn about the nature and the causes of that phenomenon. We shall not deal here with the base rate fallacy, because the latter hinges on the additional claim that the problem posed to the subjects has only one answer. Such a claim is, at best, controversial, and competing statistical schools have since time immemorial offered competing answers to the same "real-world" problems (Gigerenzer & Murray, 1987). We shall use here the term *base rate neglect* to denote the phenomenon that subjects do not use the base rates *specified* in a problem when making probability revisions, whether these base rates are considered to be "normative" or not.

The central question has become *when* and *why* base rates are neglected, rather than whether they are. Originally, base rate neglect was explained by a representativeness heuristic, conceived as a general, all-purpose heuristic for probability revision problems and beyond, even "in the evaluation of

This research was supported by Grant 179 005 85 from the University of Constance and an Akademie Stipendium from the Volkswagen Foundation to the first author. We thank Peter Sedlmeier for his assistance in data analysis.

Correspondence concerning this article should be addressed to Gerd Gigerenzer, Fachgruppe Psychologie, University of Constance, Postfach 5560, 7750 Konstanz, West Germany.

uncertainty in essentially unique situations where no 'correct' answer is available" (Kahneman & Tversky, 1972, p. 451). Meanwhile, conditions have been identified that seem to encourage the use of base rates, and although the effect of the conditions is often reported as weak, this has led to speculations about additional or alternative explanations. The first is the conjecture that base rates are used to the degree they are perceived to be "causally" relevant (Ajzen, 1977; Tversky & Kahneman, 1982), a conjecture recently modified by the assertion that only causally nonredundant base rates are used (Locksley & Stangor, 1984). The second conjecture is that base rates are neglected insofar as they are remote, pallid and abstract, as opposed to the vividness, salience, and concreteness of individuating information (Nisbett & Borgida, 1975; Nisbett & Ross, 1980). The third conjecture is that properties of the task—called "informational factors" by Ginosar and Trope (1980) and "procedural variables" by Tversky and Kahneman (1982)—influence the degree of base rate neglect.

We shall assume as a working hypothesis that the use of base rates is largely determined by the subjects' internal problem representation rather than by a general representativeness heuristic (or, equivalently, by general conservatism). Whether a problem is represented as a Bayesian-type probability revision problem depends on (a) the presentation of information (i.e., the above-mentioned task properties) and on (b) the previous experience with the content (not the structure) of the problem. It seems that there are only few inferences in everyday life that are experienced as simple Bayesian-type inference problems, where known base rates are revised, given new information. For instance, if one meets at the dean's birthday party a 50-year-old man and is asked to guess his profession, it is not a good strategy to start with the base rates of males in all professions in this country, because one must assume that this person has not been randomly drawn from the population to which the base rates refer. In Experiment 1 we shall use Kahneman and Tversky's (1973) profession problem, but let the subjects themselves do the random sampling of descriptions and thereby generate an internal representation of Kahneman and Tversky's problem as a probability revision problem. In Experiment 2, we shall use one of the rare examples in which a simple, everyday problem should already be internally represented as a Bayesian-type probability revision problem. If our working hypothesis holds, then base rate neglect should be reduced, or even disappear, in such cases.

Experiment 1: Engineer–Lawyer Problem

One of the classical demonstrations of base rate neglect is the engineer–lawyer problem (Kahneman & Tversky, 1973). Subjects were told that a panel of psychologists had written personality descriptions of 30 engineers and 70 lawyers and that five descriptions had been chosen randomly from them. For each description, the subjects had to assess the probability that the person described was an engineer. A second group of subjects was given the same descriptions but with inverted base rates—that is, 70 engineers and 30 lawyers.

The authors concluded that base rates were largely ignored because mean judgments were only 5% higher (on the probability scale) in the group with a higher base rate for engineers. This base rate neglect was explained by postulating a general heuristic called representativeness. According to the representativeness hypothesis, people judge the posterior probability solely by the similarity between the description and their stereotype of an engineer. A major support for the representativeness heuristic seemed to be the fact that the median probability judgments for the description of "Dick" were the same (.50) in both base rate groups, although this description was constructed to be completely uninformative. Subsequent studies (Ginosar & Trope, 1980; Wells & Harvey, 1978; Zukier & Pepitone, 1984), however, found that the degree of base rate neglect for the uninformative description depended on the presentation of information and the instructions.

Random sampling of a description from a population (for which the base rates are characteristic) is an essential condition for the use of the base rates specified. If the descriptions were not sampled randomly, these base rates were irrelevant. In all studies quoted above, subjects were verbally assured that all descriptions had been randomly sampled, although these were in fact deliberately constructed to match the American stereotype of an engineer or a lawyer, or to be undiagnostic with respect to both. The general question is whether verbal assertions of random sampling are sufficient to direct attention to this crucial information and to activate a corresponding problem representation. In the context of the engineer–lawyer problem, the additional question emerges of whether a verbal assertion of random sampling is effective if it is not true: There might be cues indicating the latter to the subjects—for instance, the extremely stereotyped descriptions. If the subjects were suspicious, they might be better off to ignore the specified base rates, but we do not know, because the subjects were not asked.

In this experiment we wanted to see what happens if the subjects actually observed the random sampling process. Experiment 1 consisted of (a) a straightforward replication of the engineer–lawyer study of Kahneman and Tversky (1973), using a verbal assertion of random sampling, and (b) a modification with visually observable random sampling. Witnessing the sampling process, we expected, would focus attention on the base rates and induce a problem representation where base rates are a relevant part. We expected base rate neglect to disappear.

Method

Subjects

Subjects were 97 male and female students at the University of Constance who were paid for participating in the experiment. Subjects were tested individually in the visual observation condition and in small groups in the verbal assertion condition. Subjects had no knowledge about studies on base rate neglect.

Materials

Six personality descriptions were used. From a sample of real descriptions of engineers and lawyers provided by their close friends, four were selected. In addition, two hypothetical descriptions from

Kahneman and Tversky (1973) were used. These included the description of Jack, an American stereotype of an engineer, and of Dick, the so-called uninformative description. The real descriptions contained one of an engineer and three of a lawyer. All real descriptions were of about the same length as the description of Jack. One of the real descriptions follows:

> Thomas O. is 30 years old and single. He seems a bit quiet, but in social situations is sure of himself and personable. He is reliable. Success is not the most important thing in life for him. He is interested in local politics, and is involved in environmental issues. His hobby is jogging.

In the verbal assertion condition, subjects were given the descriptions and the instruction in a questionnaire. In the visual observation condition, small sheets of paper were used which had on one side a description and on the other side a capital letter, either *E* for engineer, or *L* for lawyer.

Design and Procedure

This was a 2 × 2 between-subjects design with base rate of engineers (30% vs. 70%) and reality of sampling process (verbally asserted vs. visually observed) as the independent variables. Subjects were randomly assigned to the four conditions. The dependent variable was the rated posterior probability that the described person was an engineer.

The two verbal assertion groups are a straightforward replication of Kahneman and Tversky's (1973) engineer–lawyer experiment. The original cover story given to our subjects is repeated for convenience here:

> A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms six descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100. The same task has been performed by a panel of experts, who were highly accurate in assigning probabilities to the various descriptions. You will be paid a bonus to the extent that your estimates come close to those of the expert panel.

These instructions were given to a group of 24 subjects (low base rate/questionnaire group); a second group of 24 subjects (high base rate/questionnaire group) was given identical instructions except that they were told that there were 70 engineers and 30 lawyers. After the cover story, all subjects received the same six descriptions, and they estimated for each the probability that the described person was an engineer.

The order of presentation of the six descriptions was systematically varied to control for ordering effects; that is, in each group each description was given to 4 subjects as their first description to judge. Following the six descriptions, all subjects encountered the *null* description: "Suppose now that you are given no information whatsoever about a person chosen at random from the sample. The probability that this man is one of the 30 engineers in the sample of 100 is __ %." In the two visual observation groups (*N*s = 24 and 25) the same procedure was followed except that the randomness of sampling was made visually observable. The subject was shown 10 sheets of paper, 3 of them (or 7, depending on the base rate condition) marked with an *E* for engineer and 7 (3) marked with an *L* for lawyer. The subject could not read the description on the other side. Then the experimenter folded the sheets so that neither the identification letters nor the description could be read, threw them into an empty urn, and shook them. The subject drew one description randomly from the urn and gave it to the experimenter, who unfolded it in a way that the subject now could read the description but not the identification. After having observed the random drawing from an urn, the subject judged the probability that the person described was an engineer.

Each subject made six drawings, each from a different urn. The way to ensure that all subjects judged all six descriptions (and to avoid repeated drawings of the same description) was to have in each urn the same description on every sheet. This was completely unnoticeable by the subjects, because there was only one drawing from each of the six urns.

After the experiment, all subjects were given a short questionnaire, in which they were asked the following questions: (a) How did you arrive at your probability judgments? Try to describe this process as precisely as possible. (b) When (under what condition) did you consider the base rates of engineers and lawyers?

Results and Discussion

Results are summarized in Table 1. The first two columns show the mean probabilities in the verbal assertion group that the person described was an engineer, given a base rate of 30% and 70% engineers, respectively. All means were larger in the 70% condition, and the overall difference, averaged across all six descriptions, was 9.8% on the probability scale. For comparison, Kahneman and Tversky (1973) reported an overall difference of 5%, which, although significant, they considered small enough to justify the conclusion that base rates were largely neglected. Columns 3 and 4 show how far the observed means deviate from what base rate neglect and Bayesian performance predict, respectively. If subjects judged the probabilities by a representativeness heuristic, then base rates should be neglected, and the difference between the means in the two base rate conditions (columns 1 and 2) should be zero. Column 3 contains the differences between columns 1 and 2 and shows the deviations of actual judgments from base rate neglect. In deriving the predictions in column 4, we followed Wells and Harvey (1978), who have shown that earlier calculations of Bayesian predictions for the present design were suboptimal. Predictions for Bayesian performance were derived in the following way. For each subject in the 70% base rate group we predicted, by using Bayes's theorem, how this subject would have responded in the 30% base rate group. This means that from the distribution of the actual judgments (p) in one base rate group we predicted the corresponding distribution of judgments ($\hat{p}$) in the other group. The mean of all $\hat{p}$ scores in each base rate group could thus be compared with the obtained mean in the other base rate group. Column 4 shows the difference between the obtained means in the 30% group and the predicted means for

Table 1
Mean Judgments of the Probability That the Person Described Is an Engineer (in %) and Deviations From Predictions

	RS verbal				RS visual			
	Means		Dev. from pred.		Means		Dev. from pred.	
Description	30%	70%	BRN[a]	Bayesian[b]	30%	70%	BRN[a]	Bayesian[b]
Thomas	45.8	62.5	16.7	16.7	46.0	69.0	23.0	6.6
Peter	44.9	60.6	15.7	15.8	39.2	64.8	25.6	6.5
Heinz	34.0	36.3	2.3	19.9	31.4	45.8	14.4	10.6
Jack	71.4	81.3	9.9	16.1	64.4	79.5	15.1	7.8
Klaus	24.9	37.9	13.0	11.4	26.4	34.5	8.1	15.5
Uninformative[c]	60.0	61.2	1.2	30.3	55.6	73.4	17.8	13.8
Null	28.0	68.8	40.8	−3.8	28.6	69.9	41.3	−1.4
Average (sans null)	46.8	56.6	9.8	18.4	43.9	61.2	17.3	10.1

Note. RS verbal = random sampling verbally asserted; RS visual = random sampling visually observed; BRN = base rate neglect; dev. from pred. = deviation from prediction.
[a] Entries are differences between the means in the 70% and 30% base rate conditions.
[b] Entries are differences between means (base rate 30%) and Bayesian predictions derived from the distribution of judgments in the 70% base rate group (see text).
[c] Dick.

the 30% condition based on the 70% group.[1] The average difference between observed and predicted means is 18.4. (We also calculated the predictions for the 70% group on the basis of the obtained probability judgments in the 30% group, which gave similar results under all conditions). To summarize, in the replication condition with real descriptions, we find less base rate neglect: The effect of base rates is roughly twice as large as in Kahneman and Tversky's original study. However, the average deviation of observed means from Bayesian performance is still almost twice as large as from base rate neglect.

This descriptive analysis is supported by statistical analysis. For five of the descriptions, observed means deviate significantly from Bayesian performance, five $ts(46) > 2.3$, five $ps < .025$; in two cases they also deviate significantly from base rate neglect.

Columns 5 to 8 show the results when random sampling could be performed and observed by the subjects themselves. Let us take the personality description of Thomas as an example. When random sampling was only verbally asserted, mean judgments were exactly between the predictions of the representativeness hypothesis (base rate neglect) and Bayesian performance. When random sampling of descriptions could be observed, the deviation from the representativeness hypothesis increased to 23.0 points (on the probability scale). The deviation from Bayesian performance decreased to 6.6 points, which is a nonsignificant deviation. For all six descriptions, the deviation from base rate neglect increased on the average from 9.8 to 17.3, and the deviation from Bayesian performance decreased from 18.4 to 10.1. This shows that when random sampling could be visually observed, mean probability judgments were closer to Bayesian predictions than to the predictions of the representativeness hypothesis. If one imagines subjects' mean judgments as a point on a continuum between both predictions, with base rate neglect as zero and Bayesian performance as 1, the degree of base rate use can be expressed numerically. We define an index b of base rate use as follows:

$$b = \frac{\text{deviation from base rate neglect}}{\text{deviation from base rate neglect} + \text{deviation from Bayesian performance}}. \quad (1)$$

On this scale, the degree of base rate use is $b = .35$ in the verbal assertion condition and .63 in the visual observation condition.

Again, statistical analysis supported these descriptive results. The number of significant deviations from Bayesian performance decreased to one in the visual observation condition, and, similarly, the overall deviation from base rate neglect was significantly increased in the visual observation group, $t(5) = 2.56$, $p < .05$.

The Uninformative Description

An interesting fact in Kahneman and Tversky's original study was that the median probability judgments of Dick's being an engineer were the same (.50) in both base rate groups, although the description was constructed to be completely

[1] In formal terms, the predictions for Bayesian performance were derived as follows. From a subject's judgment $p_{70}(E|D)$ that the person is an engineer, given the description D and the 70% base rate, we calculated $\hat{p}_{30}(E|D)$ (which is what that subject would have responded, given D and the 30% base rate, assuming Bayesian performance) by using the following equations:

$$p_{70}(E|D)/(1\text{-}p_{70}(E|D)) = L\ 0.7/0.3, \quad (1)$$
$$\hat{p}_{30}(E|D)/1\text{-}\hat{p}_{30}(E|D) = L\ 0.3/0.7, \quad (2)$$

where L is the likelihood ratio (see Equations 2 and 3 below), and 0.7 and 0.3 are the base rates. Dividing Equation 1 by Equation 2 cancels out L, and $\hat{p}_{30}(E|D)$ can be calculated as a function of $p_{70}(E|D)$. The comparison of the mean of the predicted values $\hat{p}_{30}(E|D)$ with the mean of the observed values $p_{30}(E|D)$ avoids a fundamental flaw which would result from the simpler method of comparing the means of the observed values in both base rate groups: Even if subjects in both groups conformed to Bayes' theorem, their means usually would not, because Bayes' theorem predicts a curvilinear relation between both groups (see Wells & Harvey, 1978).

uninformative. In a subsequent study, however, Wells and Harvey (1978) found a substantial mean difference (.54 and .36). Ginosar and Trope (1980), who varied base rates and diagnostic information between subjects (i.e., gave only one description and base rate to each subject), found an even larger difference. In the verbal assertion condition, we could replicate the original finding of Kahneman and Tversky: Mean ratings of Dick were practically the same in both base rate groups. However, when the random sampling process was made visually observable, the difference increased from 1.2 to 17.8 (see Table 1), and base rate neglect vanished. Judgments were even closer to the Bayesian prediction, deviating significantly from the representativeness hypothesis. As measured by b, base rate use increased from .004 to .56.

How are these different results to be reconciled? We suggest that another variable of information presentation—order of descriptions—may be considered as an explanation. If a subject encounters a stereotypical description first, this may draw attention to the particular features such as the person's hobbies and his character traits, and may induce a problem representation that continues to focus attention to such individuating features in the following descriptions. That problem representations and strategies which were adopted in the first task tend to be stable in subsequent similar tasks is reflected in concepts such as "constraint" (Gebundenheit), "functional value" of units of information and "set" (Ach, 1905; Duncker, 1935/1945). We assume that such a set (to focus attention on political opinions, hobbies, etc. and to ignore base rates) can be generated by stereotypical descriptions unless the relevance of base rates is emphasized by visual observation.

None of the studies using the engineer–lawyer problem (Ginosar & Trope, 1980; Kahneman & Tversky, 1973; Wells & Harvey, 1978; Zukier & Pepitone, 1984) have analyzed the effect of the position of the uninformative description. We, too, have not designed our study as a test for order effects but have systematically varied the order; thus we could perform at least a post hoc analysis. If order is a crucial variable that accounts for the contradicting results in the literature, then we must predict that (a) in the verbal assertion condition, base rates are used if the uninformative description is encountered first but not if encountered on the second or a later position, and (b) in the visual observation condition, this interaction between position and base rate group should disappear, because attention was already directed to base rates by visualization. An investigation of the judgments of Dick showed exactly this triple interaction (see Figure 1), although it was not significant, $F(1, 89) = 3.02$, $p = .086$. The latter is probably due to the small power of this post hoc triple-interaction test, because we have only 4 individuals in each of the four conditions where Dick was presented as the first description. We may therefore tentatively propose that the order of presentation is a crucial variable for the evaluation of undiagnostic information. This could provide an explanation for the apparently contradictory results in the literature. If we are right, then given that order was *not* analyzed in all studies quoted above, the *number* of descriptions presented to the same individual is the crucial variable. The more diagnostic descriptions were used in addition to the undiagnostic one, the less frequent the undiagnostic description was the first one presented and the more base rate neglect

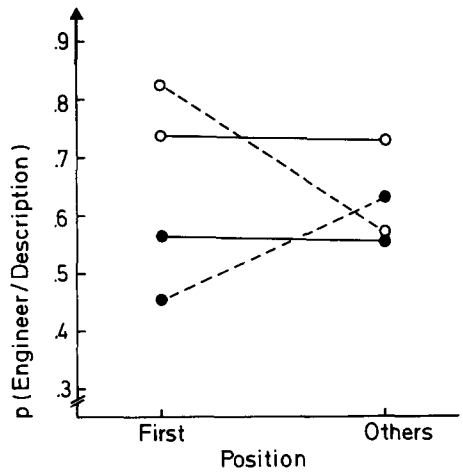

Figure 1. Order effects when random sampling is not observable. (Straight lines and dashed lines represent visual observation and verbal assertion conditions, respectively. Open and filled circles represent 70% and 30% base rate conditions, respectively).

should result for the undiagnostic information, *averaged across positions.* Kahneman and Tversky (1973) presented five descriptions, and therefore (assuming they varied the order systematically) the undiagnostic information occurred at the first position in only 20% of the cases, which predicts relatively small deviations from base rate neglect. We used six descriptions and obtained the same result. Ginosar and Trope (1980), however, presented only one description to each subject, which predicts no base rate neglect, because the undiagnostic information was always in the first position. In fact, they found a difference between base rate groups almost as large as in Figure 1 for the corresponding condition, namely, first position and verbal assertion group. Wells and Harvey (1978) and Zukier and Pepitone (1984) used one diagnostic description (Jack) in addition to the undiagnostic one. The latter report that they systematically varied the order of the two descriptions presented to each subject, but neither of the two studies analyzed the order effect. Assuming that in both studies order was varied systematically, the undiagnostic description was given first to 50% of the subjects. Thus we should expect a degree of base rate use that is between that of Kahneman and Tversky's study and Ginosar and Trope's, with 20% and 100% first positions, respectively. In fact, Wells and Harvey found a substantial mean difference, as reported above, which is smaller than that found in Ginosar and Trope's study. Zukier and Pepitone's finding is not directly numerically comparable, because they used only one base rate (.3) for all subjects. However, their finding that probability judgments for the uninformative description being an engineer ranged between .31 and .47 also shows that there is no base rate neglect if only two descriptions were used and results are averaged across positions.

The Effect of Less Stereotypical Descriptions

One question that remains is why our replication of Kahneman and Tversky's (1973) study on the engineer–lawyer

problem, using the same questionnaire method, provided an effect of base rates that was about twice as large as in the original study. The two reasons we see for this can be reduced to a single explanation. First, the real descriptions used in our study were less stereotypical than the artificially constructed ones in Kahneman and Tversky's study. Second, in Germany the stereotype of an engineer is closer to that of a lawyer than in America. Both have in common the fact that less extreme probability judgments result. As can be seen from Table 1, the average probability judgments of being an engineer did not exceed .82 and did not go below .24 (in both base rate groups). From the corresponding figure in Kahneman and Tversky (1973), we see that three of their four diagnostic descriptions result in median probabilities far beyond these figures. The point is that base rate neglect and Bayesian performance can be least differentiated for extreme probability judgments (see, e.g., Figure 1 in Kahneman & Tversky, 1973). If judged probabilities are 0 or 1, then both predictions coincide. That is, with extremely stereotypical descriptions, one cannot expect deviations from the base rate neglect curve as great as with less stereotypical descriptions. A numerical example may illustrate the size of this effect (values are chosen to represent the typical Bayesian posteriors in Kahneman and Tversky's original study and in our replication, respectively): Assume that for less stereotyped descriptions the subjective likelihoods are $p(\text{description} \mid \text{engineer}) = .7$ and $p(\text{description} \mid \text{lawyer}) = .4$. For a base rate with 70% engineers, we calculate $p(\text{engineer} \mid \text{description}) = .803$, whereas with 30% engineers we get $p(\text{engineer} \mid \text{description}) = .429$. Now assume we had used an extremely stereotypical description with $p(\text{description} \mid \text{engineer}) = .95$ and $p(\text{description} \mid \text{lawyer}) = .03$. For base rates of 70% and 30% engineers, we calculate $p(\text{engineer} \mid \text{description})$ as .987 and .931, respectively. Thus, the effect (i.e., the difference between the two base rate conditions) that can be found if subjects showed Bayesian performance is .374 on the probability scale for the less stereotyped descriptions, but only .056 for the stereotyped descriptions. Therefore, (a) natural descriptions, being less stereotyped, provide a more sensitive test whether subjects use base rates or not, and (b) Kahneman and Tversky's finding of an average base rate effect that is only half of that found in our replication can be fully explained by their use of stereotypical rather than natural descriptions.

Subjects' Reported Strategies

Subjects' reports revealed clear-cut individual differences in strategies. The five main strategies, classified by two independent observers, were the following: (a) *base rates only*—that is, subjects reported that they considered only the base rate information as relevant; (b) *qualitative Bayesian (revision type)*—that is, subjects reported that they took the base rate as their starting point and revised it according to the impact of the description, which is a strategy equivalent to Bayesian probability revision without specifying the quantitative relationships; (c) *qualitative Bayesian (integration type)*—that is, subjects reported that they integrated both base rates and diagnostic information, but without the revision idea where base rates are considered first; (d) *lexicographic representativeness*—that is, subjects reported that they evaluated the similarity between description and their mental images of engineers and lawyers first, and only if this did not differentiate enough did they also consider the base rates; and (e) *representativeness heuristic*—that is, subjects reported that they used the similarity between description and their mental images of engineers and lawyers only, as postulated in Kahneman and Tversky's representativeness heuristic. Representativeness strategies differed in content; there were subjects who reported looking only at the character description, whereas others considered mainly political and social attitudes.

Table 2 shows the percentage of subjects reporting each strategy and the mean probability ratings for individuals reporting different strategies. If reported strategies were consistent with actual judgments, then the base rate effect (difference between 30% and 70% group) should be largest for those subjects reporting to use base rates only (left side in Table 2) and should decrease from left to right in Table 2. In fact, individuals who report the base-rate-only strategy have mean values of 70.0 and 30.0, depending on the base rate group, and with increasing use of the diagnostic information the values in the two base rate groups converge. Thus, the first result is that the reported strategies are well calibrated to the actual mean probability judgments. Second, individual differences in reported strategies account for large differences in probability judgments between base rate groups, ranging from 40% for base rate only to 5.7% and 8.4% for representativeness. Finally, Table 2 shows that more individuals in the

Table 2
Subjects' Reported Strategies in the Engineer–Lawyer Problem

Statistic	BR only	Qualitative Bayesian		Representativeness		Unclassifiable
		Revision	Integration	Lexicographic	Heuristic	
Mean probability						
Base rate 70%	70.0	68.0	60.9	55.4	54.1	
Base rate 30%	30.0	35.6	43.7	49.7	45.7	
Percentage of individuals						
Verbal group ($N = 48$)	2.1	12.5	9.4	41.7	26.0	8.3
Visual group ($N = 49$)	7.1	14.3	17.3	40.8	14.3	6.1

Note. BR = base rates. Strategies were classified by two independent raters. Raters agreed in 78.4% of their classifications. With few exceptions, different classifications were restricted to those *within* the two qualitative Bayesian strategies or *within* the two representativeness strategies. Percentages reported are averaged across raters.

verbal assertion group report the use of a representativeness heuristic than in the visual observation group. In both groups, however, the lexicographic representativeness strategy is the most frequent one reported. In other contexts, lexicographical models of judgment have been proposed by Tversky (1969) and Montgomery (1977). The basic idea of lexicographical judgment is that people choose between two (or more) alternatives on the basis of a single piece of information, and only if this looks inconclusive do they consider a second kind of information. If subjects judge this way, then the degree of stereotypicality of the descriptions used is a crucial variable. In fact, this explains apparently contradictory results. For instance, the lexicographic strategy predicts that Kahneman and Tversky's (1973) subjects, who were given highly stereotypical descriptions, would largely neglect base rates, whereas Ginosar and Trope's (1980) subjects, who were given inconsistent descriptions, and our subjects, who were given real descriptions, would not, as was found.

To conclude, we have shown that random sampling, an essential condition for the applicability of Bayes's theorem (to the base rates specified), is a crucial informational variable for base rate use by our subjects. If attention is directed to random sampling by the presentation of information (e.g., visual observation), performance is closer to Bayes's theorem than to base rate neglect.

Experiment 2: Soccer Problem

There are two features of the engineer–lawyer problem that we shall single out for further investigation: subjects' familiarity with this type of profession guessing and the issue of intraindividual variation of base rates. First, our subjects were not familiar with guessing the profession of a person from a brief description and, in particular, not with interpreting such a problem in the sense of a Bayesian-type probability revision problem. If they did happen to guess professions in everyday life, then the target people were not randomly drawn from a population. For instance, there is a famous German TV program in which experts guess the profession of a candidate, who answers only yes or no to their questions. Here, the main heuristic strategy is searching for further information, whereas base rates are not relevant, because the candidates were selected and not randomly drawn. In fact, the experts would perform badly if they started with the base rates of professions and revised these according to Bayes's theorem; rather, they must anticipate the rarest professions.

Second, base rate information was never varied intraindividually, whereas descriptions were. In fact, almost all the experiments from which base-rate neglect (or strong underweighting) was concluded presented different diagnostic information to each subject but varied the base rates only between subjects. Thus, every subject had different diagnostic information, but no one had different base rates. Of course, the engineer–lawyer problem is not suitable for varying both base rates and diagnostic informations to the subjects, because this would mean instructing the subject that there were, say, four populations of engineers and lawyers, with different proportions of engineers, all interviewed and described by psychologists, and from each population some persons were randomly drawn, and so on; that would make the problem look very artificial, to say the least.[2]

To investigate the effect of familiarity and intraindividual variation of diagnostic information, we selected a problem where probability revision is familiar from everyday life and where varying base rates is as natural as varying diagnostic information: the soccer problem.

Every week, millions of spectators watch the games of the German soccer "Bundesliga," in which the 18 best teams compete. Every week, the spectators' situation is similar to Bayesian probability revision. Before the game they have some expectation about the probability that Team A will win, based on the earlier performance of that team. During the game they get new diagnostic information—for example, that their favored team is two goals behind—and may revise their original probability in light of this new information. Although spectators may not always specify the prior probabilities of a win quantitatively, this everyday situation has a clear parallel to Bayesian probability revision.

Varying base rates comes as naturally as varying diagnostic information, because the proportion of games won in a season varies from team to team and from year to year. We used a team's number of won games in a season as the base rate, and the halftime result as the diagnostic information. The subject's task was to estimate the probability that Team A won a game, randomly chosen from a season, given Team A's proportion of wins at that season and the halftime result of the game. Thus the subject's task had the same structure as in the engineer–lawyer problem: to revise a prior probability (proportion of wins) into a posterior probability in the light of diagnostic information (the halftime result). Four base rates and three classes of halftime results (draw, one goal ahead [+1], and one goal behind [−1]) were used.

Figure 2 shows the predictions of three prominent hypotheses: representativeness, base rates only, and Bayesian performance. First, the assumption of a representativeness heuristic predicts that our subjects would judge the posterior probability of win by the similarity between an event (halftime result) and a class (that of won games). A halftime result of one goal ahead is more "similar" to the results that characterize won games than is a draw at halftime. Similarly, a draw at halftime is more "similar" to the results characterizing won games than is a halftime result of one goal behind. This hypothesis predicts parallel curves, that is, judgments that are independent of the base rates (Figure 2a). Note that the distances between the +1, draw, and −1 curves, as well as the absolute levels, are not specified by the representativeness hypothesis.

Second, what we call the base-rates-only hypothesis is an extreme version of conservatism (Edwards, 1968), and Figure 2b shows the prediction when only base rates are used.

[2] Hell, Gigerenzer, and Bodenburg (1987) performed such an experiment but used only two populations of engineers and lawyers. They found that the difference between the conditions of inter- and intraindividual variation of base rates was practically zero. This result conforms well with that of Experiment 2, reported below.

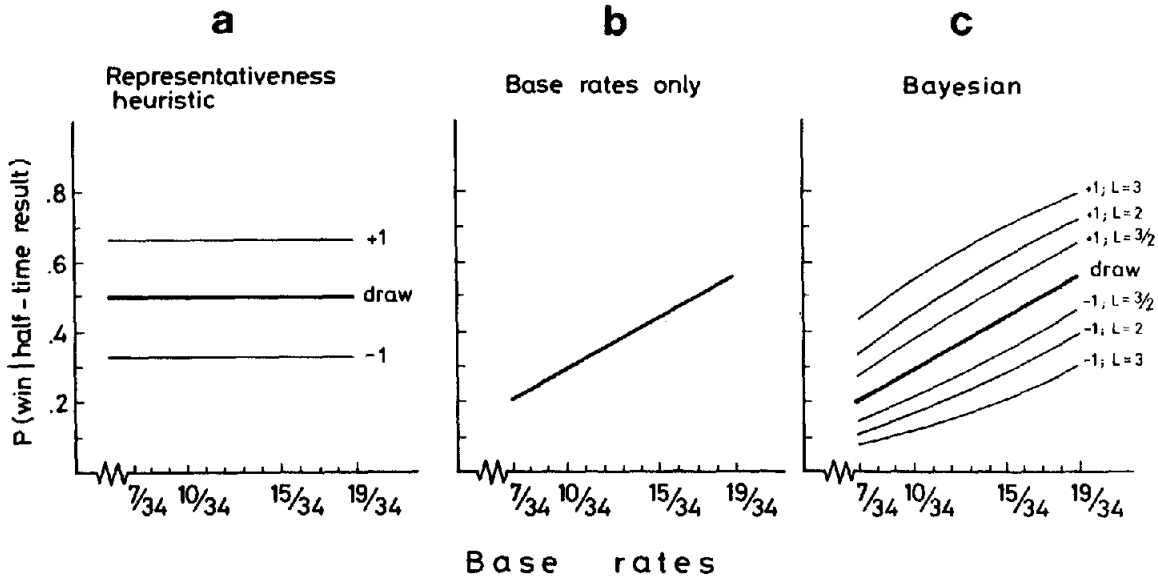

Figure 2. Predictions for the soccer problem, assuming that (a) subjects use a representativeness heuristic, (b) subjects use base rates only (i.e., are extremely conservative), and (c) subjects use a Bayesian strategy. (Predictions for [c] depend on the likelihood ratio L; three values of L are given for illustration [see text]. Base rates specify the number of won games out of a total of 34 games played in a season).

Third, Bayesian predictions can be derived from the odds version of Bayes's theorem (using +1 as an example):

$$\frac{p(\text{win}\,|\,+1)}{p(\text{no win}\,|\,+1)} = \frac{p(\text{win})}{p(\text{no win})} \times \frac{p(+1\,|\,\text{win})}{p(+1\,|\,\text{no win})}. \qquad (2)$$

What the subjects judge is $p(\text{win}\,|\,+1)$, that is, the probability that the game is one of the won games, given the information about the halftime result; $p(\text{no win}\,|\,+1)$ is the complement, and the base rates $p(\text{win})$ and $p(\text{no win})$ are specified. What is unknown is the likelihood ratio $p(+1\,|\,\text{win})/p(+1\,|\,\text{no win})$. This likelihood ratio determines the "distance" between the three curves. To illustrate possible Bayesian predictions to the soccer problem, we shall use two simplifying assumptions.

First, we assume that our subjects consider a draw at halftime as an undiagnostic information. Therefore, the draw curve in Figure 2c is the base rate curve from Figure 2b. The second simplifying assumption is that the +1 and −1 curves are "symmetric" around the draw curve, that is, that the following two likelihood ratios are equal:

$$L = \frac{p(+1\,|\,\text{win})}{p(+1\,|\,\text{no win})} = \frac{p(-1\,|\,\text{no win})}{p(-1\,|\,\text{win})}. \qquad (3)$$

We illustrate in Figure 2c the Bayesian prediction of the soccer problem for three simple likelihood ratios, $L = 3/1$, $2/1$, and $3/2$.[3] For instance, a ratio of 2/1 means that the subjects believe that the probability of being one goal ahead in won games is twice as high as in games not won.

This is the most simplified Bayesian model of the soccer problem we can think of; for that reason we shall start with this model. If subjects' judgments were Bayesian, however, with nonsymmetric likelihood ratios, we would have to turn to less restricted Bayesian models. An example would be the combination of the $L = 3/1$ curve for one goal ahead and $L = 2/1$ curve for one goal behind (see Figure 2c).

In terms of analysis of variance, neglect of base rates predicts that the halftime results should show the only significant effect. Extreme conservatism predicts that only the main effect for base rates is significant, whereas all Bayesian models (including the simplified one) predict that both main effects and the interaction are significant (because the Bayesian lines are curved, not parallel, and diverge from left to right, as can be seen from Figure 2c).

Recall that our working hypothesis was that the internal problem representation, rather than a general representativeness heuristic, directs probability judgments. If this is true, then the familiarity of the soccer problem as a probability revision problem and the intraindividual variation of base rates should generate a problem representation that nullifies base rate neglect. In order to control for the separate impacts of familiarity and intraindividual base rate variation, we have performed two parallel experiments, using the soccer problem. In the first, base rates and diagnostic information were varied in a within-subjects design; in the second, diagnostic information was varied in a within-subjects design, but base rates were varied between subjects, as in the engineer–lawyer study.

[3] As in Figure 2a and 2b, predictions are given in terms of $p(\text{win}\,|\,\text{halftime result})$ rather than in terms of odds, as in Equation 2. The relation between the two formats is the following. If we rewrite Equation 2 as $\Omega_1 = \Omega_0\, L$, with Ω_1 and Ω_0 as the posterior and prior odds, respectively, then $p(\text{win}\,|\,\text{halftime result}) = \Omega_1/(1 + \Omega_1)$. The reason why the Bayesian lines do not reconverge at the upper end of Figure 2c is that the range of base rates used is not as extreme on the upper end as it is on the lower.

Method

Intraindividual Variation of Base Rates

Subjects. Forty-eight male and female students were paid for participating in the experiment. Subjects were tested in group sessions.

Design and procedure. All soccer problems were of the following type:

> In the 1978/79 season of the West-German Soccer "Bundesliga," Team A won 10 out of the 34 games. The other games were either drawn or lost. We have selected some of the games of that season randomly and checked their final results as well as their halftime results. For instance, on the 7th day of the season the halftime result was 2:1 in favor of Team A. What is your probability that this game belongs to those 10 games won out of 34?

The design was a 3 × 4 within-subjects design, with three classes of halftime results (+1; draw; −1) and four base rates (7, 10, 15, or 19 games won out of 34). Draws included 0:0, 1:1, and 2:2, +1 included 1:0 and 2:1; −1 included 0:1 and 1:2. Before encountering these 12 soccer problems, each subject was given four other soccer problems, which were not analyzed and which served to get the subjects acquainted with that type of problem. The order of the 12 problems was randomly determined.

After the subjects had completed the soccer problems, they were given a questionnaire with the same kind of questions as in the engineer–lawyer problem. The following question was added: "Over all, what was more important for you—the halftime results or the number of games won? Please try to express this in percentages!"

Interindividual Variation of Base Rates

Subjects. One hundred forty-four male and female students participated in group sessions.

Design and procedure. The design and procedure were same as before, except that each individual received a set of three soccer problems (as opposed to 3 × 4), all with the same base rate but with three different halftime results. Thus, a quarter of the subjects received each base rate.

Results and Discussion

Variation of Base Rates Within Subjects

Figure 3 (left side) presents the median probabilities of the game's being a win for the four base rates and three classes of diagnostic information. The medians reject both the base-rate neglect and the extreme conservatism hypotheses but can be fitted to the simplified Bayesian model, with a likelihood ratio of 2. Of course, a better fit to a Bayesian prediction could be obtained if we relaxed the strong symmetry assumption (Equation 3), which would allow increasing the values for the −1 curve (without decreasing those for the +1 curve). But we see no gain in this, because the simplified version already gives an excellent fit.

An analysis of variance using the means shows the same result. The hypothesis that judgments in the four base rate groups stem from the same population is rejected, $F(3, 141) = 108.5$, $p < .001$; in addition, judgments vary significantly with the diagnostic information, $F(2, 94) = 66.6$, $p < .001$; finally, the interaction is also significant, $F(6, 282) = 4.3$, $p < .001$, as predicted by the Bayes rule. The size of effect can be seen from Figure 3; the maximum differences due to base rates and diagnostic information are .45 and .35 on the probability scale, respectively. This, as well as the good Bayesian fit of the data, rules out both base rate neglect and

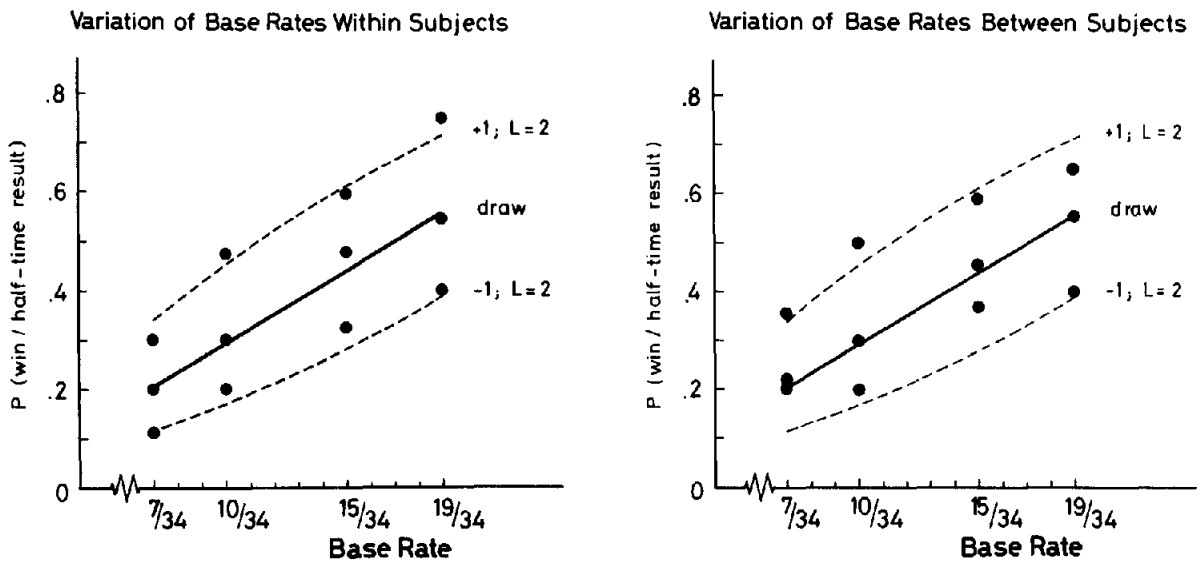

Figure 3. Median probability judgments using a within-subjects and a between-subjects design for variation of base rate information. (Data points are fitted by a simplistic Bayesian model with a likelihood ratio $L = 2$ [see text]. The solid line shows the predicted probabilities for a draw; the dashed lines show those for one goal ahead [+1] and one goal behind [−1].)

conservatism, including all weak versions thereof, such as the claim that subjects "largely" ignore base rates.

However, there is some danger from drawing conclusions from means or medians to the performance of individual minds. For instance, if half of our subjects were to use a representativeness heuristic and the others were extremely conservative, then means would produce curves similar to those in Figure 3. An inspection of all individual data plots showed no sign of such difference between subjects (there was only 1 subject who ignored the halftime results, but none ignored the base rates). A more formal test was performed by a conjoint measurement analysis of single cases (Falmagne, 1985), which showed that 31% of our subjects satisfied the testable necessary conditions (independence, double cancellation) for an additive representation perfectly, and almost all others showed only a few unsystematic violations. Because Bayes's rule is a special case of the additive conjoint model, this single case analysis supports that of the medians and means.

Variation of Base Rates Between Subjects

Figure 3 (right side) shows that even when no varying base rate information is given to a subject, there is no base rate neglect in the soccer problem. The medians for the draw results coincide again in three cases exactly with the predictions, indicating that a draw is considered as undiagnostic information. The medians for one goal ahead can be fitted again by a likelihood ratio of 2, but the medians for one goal behind show a more pronounced deviation from the simplified symmetry assumption. Here, judgments may be fitted by a likelihood ratio around 3/2 (see Figure 2c). It indicates that subjects believe that the probability of being one goal ahead in won games is twice as high as in games not won but that the probability of being one goal behind in games not won is only 1.5 times as high as in won games. They seem to consider the diagnostic impact of a positive information (one goal ahead) to be greater than that of a negative information, which mirrors the optimistic bias of the stereotypical soccer fan: If our team is ahead, we shall certainly win, if we are behind, we shall possibly win.

Analysis of Subjects' Reports

Reported strategies for the soccer problem (intraindividual variation condition) were qualitatively the same as in the profession problem, but with a different distribution. Table 3 shows that 77% of all subjects reported strategies relying on base rates, almost all of them reporting qualitative Bayesian strategies. There is no single subject reporting a representativeness heuristic. The comparison of the frequencies in the first three strategies with the two representativeness strategies shows a significant dependency of strategies on the familiarity of the problem (soccer vs. profession) with $\chi^2(1, N = 145) = 28.0$, $p < .001$. On the average, the reported weight given to base rates was greater (59.1%) than to the diagnostic information (40.9%). These reported weights correspond well with the actual maximum differences reported earlier, .45 and .35 on the probability scale. Subjects' reports are well calibrated to their probability judgments.

External Validity

Besides the question, When do people use the base rates specified in a problem?, another question arises: Do people's judgments correspond to reality? The present experiments allow a check on the external validity of judgments, because we have used real descriptions in Experiment 1 and we can compare the results in Experiment 2 with the real facts in a soccer season. As mentioned above, the four real descriptions used in Experiment 1 were from three lawyers (Peter, Heinz, and Klaus) and one engineer (Thomas). If subjects' probability judgments were externally valid, the mean probability judgments for Thomas (and Jack, the stereotypical description) should be greater than for all others. Table 1 shows that this is the case in all four experimental groups.

For the soccer problem we can perform a more quantitatively elaborated validation. We analyzed the whole 1985/86 season in the German Bundesliga including 306 games and calculated from this the actual likelihoods. Recall that our simplified Bayesian model assumed that subjects consider a draw as undiagnostic information and that subjects' judgments conformed to this simplifying assumption. The actual likelihoods, however, were $p(\text{draw}|\text{win}) = .278$ and $p(\text{draw}|\text{no win}) = .387$. This gives a likelihood ratio of .718, which is smaller than that of 1.0—which corresponds to our subjects' judgments. This means that our subjects overestimated the probability $p(\text{draw}|\text{win})$ in relation to $p(\text{draw}|\text{no win})$. In reality, the two properties are not equal; a draw at halftime occurs less frequently in won games than otherwise. This discrepancy can be seen from Figure 4, where solid lines show the Bayesian predictions based on the real likelihoods and dashed lines represent subjects' judgments, approximated by the simplified model from Figure 3. The figure also shows that the impact of being one goal ahead or behind on the probability that a team will win the game is stronger than our subjects believed. The actual likelihoods $p(+1|\text{win})$ and $p(+1|\text{no win})$ are .378 and .123, respectively. This means that the likelihood ratio is 3.07 as opposed to 2.0, which was a good approximation to our subjects' judgments. Recall that the more the likelihood ratio deviates from 1, the more impact the diagnostic information has. The same underestimation

Table 3
Subjects' Reported Strategies

		Qualitative Bayesian		Representativeness		
Task	BR only	Revision	Integration	Lexicographic	Heuristic	Unclassifiable
Soccer problem	2.0	41.7	33.3	16.7	0.0	6.3
Profession problem (average)	4.6	13.4	13.4	41.2	20.1	7.2

Note. BR = base rates.

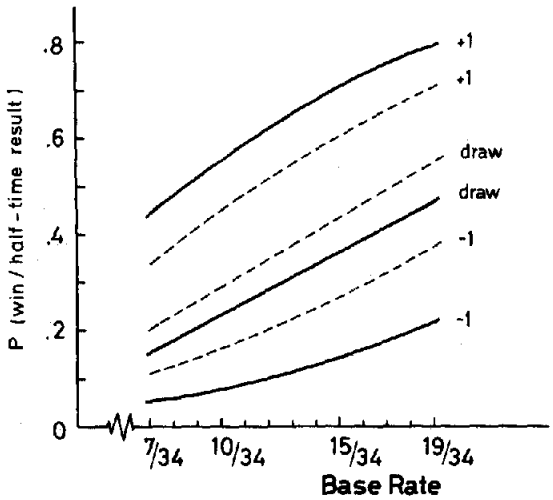

Figure 4. External validity of soccer judgments. (Solid lines are probabilities p[win | halftime result] derived from the actual games in a soccer season; dashed lines represent subjects median judgments [see Figure 3].)

occurs when a team is one goal behind. The real likelihood ratios are $p(-1 \mid \text{win}) = .070$ and $p(-1 \mid \text{no win}) = .309$. Thus, the likelihood ratio is .227, or expressed in its inverted form, as in Equation 3, the ratio is 4.41. Again, subjects did not revise the base rates enough, given the information that a team is one goal behind. Thus, although our subjects' judgments satisfied Bayesian performance, they strongly underestimated the impact of all halftime results, including a draw. Measured against the real soccer results, our subjects were conservative (e.g., Edwards, 1968). There is another aspect where both our subjects and the simplified Bayesian model deviate from reality. Recall that the symmetry assumption (Equation 3) assumes that the likelihood ratio for being one goal ahead is the same as the (inverted) likelihood ratio for being one goal behind. In reality, however, the two corresponding ratios were 3.07 and 4.41. We have noticed earlier, that our subjects' judgments also deviated slightly from the −1 curve (based on the symmetry assumption and a likelihood ratio of 2). This deviation was, however, in the opposite direction from the actual soccer games, where the diagnostic impact of being one goal behind is larger than that of being one goal ahead. Our subjects acted as if they believed that being one goal behind is less decisive than it actually is.

To summarize, compared with the soccer reality, our subjects used the diagnostic information conservatively; in particular, they seemed not to realize that a draw is not undiagnostic information, and they showed a bias toward optimism in the case of a negative halftime result. Even good Bayesians can be wrong about reality; probability theory and rationality may indeed diverge at some points.

General Discussion

We succeeded in reducing the base rate neglect phenomenon in the engineer–lawyer study by making the crucial condition of random sampling visually observable. This pushed the probability judgments more than halfway toward Bayesian performance. In the soccer problem, where we simulated everyday estimates of the probability that a team would win a game, given a particular halftime result, subjects' judgments were indistinguishable from Bayesian performance. Although the formal structure of both problems was the same, the presentation of information and the content of the problem strongly influenced both subjects' performance and their reported strategies.

Results in both experiments imply that the original explanation, the representativeness heuristic, is inadequate. We conjecture that probability judgments in problems like those studied here are to be explained by the internal problem representation induced rather than by the exercise of a general heuristic. Whether a problem is internally represented as a probability revision problem, in which base rates are considered as a relevant information, is dependent on the actual presentation of information and on the previous experience with the content. We have identified the visual observation of the random sampling process as a crucial variable of information presentation and the familiarity with a topic as one of probability revision in everyday life as a crucial variable of content. The representation of a problem may, in addition, be directed by knowledge about statistical rules such as Bayes's theorem, but this is not an important factor in the case of our subjects. Instead of searching for heuristics, understood as general, all-purpose rules which are independent of context and content, we suggest an analogy between perceptual judgment and probability judgment as a working hypothesis. The analogy concerns the investigation of variables that direct the problem representation, such as stimulus position, variation, repetition, temporal order, spatial order, and other contextual variables that define the presentation of information, direct attention, and influence perceptual judgment (e.g., Birnbaum, 1982; Hell, 1987; Parducci, 1974).

In Search of a General Perspective

Our working hypothesis was that base rate use is dependent on an internal problem representation. We consider this internal representation to be mediated by two factors, first, the presentation of information in the actual experiment, which directs the distribution of attention over the information available, and second, the previous experience with the content of the problem (Figure 5). We conjecture that base rate neglect is not a general property of the human mind and that the representativeness heuristic is not a general-purpose strategy in probability revision problems. If our hypothesis is

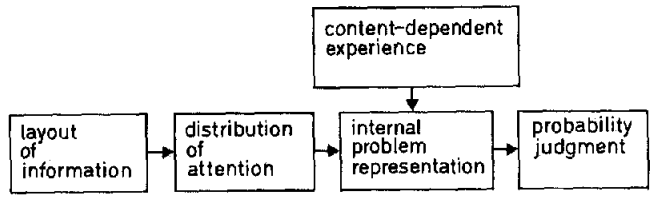

Figure 5. A schema of human probability revision as an alternative to the general cognitive heuristics view.

correct, then questionnaires and other methods of displaying a problem are not mere neutral tools but are of essential theoretical relevance. For instance, because questionnaires cannot make crucial structural components such as random sampling visible and manipulable, they will generate a problem representation different from the method used in our engineer–lawyer study. Verbal assertions alone cannot always replace visual experience.

Duncker (1935/1945, p. 1) asked the question, "How does the solution arise from the problem situation?"; however, most recent experimenters, in effect, ask the question, "How does the solution arise from a general cognitive heuristic?" Duncker's emphasis, as well as that of Ach (1905) and Messer (1906), was on the stimulus side, whereas more recent explanations seek general-purpose heuristics of the brain that are believed to operate more or less independently of the presentation of information as well as of the content of the problem. We suggest Figure 5 as a general perspective. Probability revision and base rate use is understood as dependent on external presentation of information and internal content-dependent experience. What is the relation between this perspective and the heuristics view? In order to answer, we must distinguish different versions of the heuristics view. If, in probability revision problems, representativeness is considered as a *general, all-purpose* heuristic where only structure, not content and presentation, counts (Gigerenzer & Murray, 1987, chapter 5), then the present perspective, although general, contradicts it. Moreover, this perspective also differs from the weaker interpretation that people have a *tendency* to use a representativeness heuristic in probability revision problems, because, as we suggest, presentation and content largely determine whether a representativeness judgment is made or not. We consider a representativeness heuristic a special case, one elicited by factors that are irrelevant from the standpoint of a purely structural theory. There is experimental evidence that probability judgments are influenced by the presentation of information. Examples of variables affecting probability judgments are (a) visual observation of likelihoods and base rates (e.g., Christensen-Szalanski & Beach, 1982); (b) visual observation of random sampling (Experiment 1); (c) order of presentation of problems (e.g., Duncker, 1935/1945; tentative evidence from Experiment 1); (d) repetition of information (e.g., Gigerenzer, 1984); and (e) intraindividual variation of information (Birnbaum & Mellers, 1983; tentative evidence from Fischhoff, Slovic, & Lichtenstein, 1979). Furthermore, content-dependent experience has been shown to affect base rate use (e.g., Alker & Hermann, 1971; Evans, Brooks, & Pollard, 1985).

We conjecture that given two problems with the same formal structure, the presentation of information and the content are so crucial that base rate use can be produced in one case but representativeness strategies in the other. If we are correct, then the apparently contradictory results found since the 1960s—from "conservatism" to "base rate neglect," the very antithesis of "conservatism"—are what is to be expected in studies on probability revision unless the "technical" modes of information presentation and the "irrelevant" content of the problems are controlled and incorporated into a theory of judgment under uncertainty.

References

Ach, N. (1905). Über Willenstätigkeit und das Denken [On the action of will and on thinking]. Göttingen, W. Germany: Vandenhoeck & Ruprecht.

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.

Alker, H. A., & Hermann, M. G. (1971). Are Bayesian decisions artificially intelligent? The effect of task and personality on conservatism in processing information. *Journal of Personality and Social Psychology, 19,* 31–41.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysics* (pp. 401–485).

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

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology, 45,* 792–804.

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

Daston, L. J. (1980). Probabilistic expectation and rationality in classical probability theory. *Historia Mathematica, 7,* 234–260.

Duncker, K. (1945). On problem solving (L. S. Lees, Trans.) *Psychological Monographs, 58*(5, Whole No. 270). (Original work published 1935)

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

Evans, J. St B. T., Brooks, P. G., & Pollard, P. (1985). Prior beliefs and statistical inference. *British Journal of Psychology, 76,* 469–477.

Falmagne, J. C. (1985). *Elements of psychophysical theory.* Oxford: Oxford University Press.

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

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency–validity relationship. *American Journal of Psychology, 97,* 185–195.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

Ginośar, 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.

Hell, W. (1987). *Aufmerksamkeit und Interferenz* [Attention and interference]. Weinheim, W. Germany: Beltz.

Hell, W., Gigerenzer, G., & Bodenburg, S. (1987). *Processing of base rate information: Does intraindividual variation help?* Unpublished Manuscript, University of Constance, Federal Republic of Germany.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Laplace, P. S. (1951). *A philosophical essay on probabilities.* F. W. Truscott & F. L. Emory, Trans.). New York: Dover. (Original work published 1814)

Locksley, A., & Stangor, C. (1984). Why versus how often: Causal

reasoning and the incidence of judgmental bias. *Journal of Experimental Social Psychology, 20,* 470–483.

Messer, A. (1906). Experimentell-psychologische Untersuchungen über das Denken [Experimental–psychological investigations on thinking]. *Archiv für die gesamte Psychologie, 8,* 1–224.

Montgomery, H. (1977). A study of intransitive preferences using a thinking aloud procedure. In H. Jungermann & G. DeZeeuw (Eds.), *Decision making and change in human affairs* (pp. 347–362). Dordrecht, Holland: Reidel.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32,* 932–943.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment* (pp. 127–141). Englewood Cliffs, NJ: Prentice-Hall.

Parducci, A. (1974). Contextual effects: A range frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception,* Vol. II. New York, Academic Press.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72,* 346–354.

Porter, T. (1986). *The rise of statistical thinking: 1820–1900.* Princeton, NJ: Princeton University Press.

Shafer, G. (1978). Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences, 19,* 309–370.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76,* 31–48.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge: Cambridge University Press.

Wells, G. L., & Harvey, J. H. (1978). Naive attributors' attributions and predictions: What is informative and when is an effect an effect? *Journal of Personality and Social Psychology, 36,* 483–490.

Zukier, H., & Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology, 47,* 349–360.

Received June 12, 1987
Revision received November 9, 1987
Accepted November 17, 1987 ■