

THEORETICAL NOTES

On the Reality of Cognitive Illusions

Daniel Kahneman
Princeton University

Amos Tversky
Stanford University

The study of heuristics and biases in judgment has been criticized in several publications by G. Gigerenzer, who argues that "biases are not biases" and "heuristics are meant to explain what does not exist" (1991, p. 102). This article responds to Gigerenzer's critique and shows that it misrepresents the authors' theoretical position and ignores critical evidence. Contrary to Gigerenzer's central empirical claim, judgments of frequency—not only subjective probabilities—are susceptible to large and systematic biases. A postscript responds to Gigerenzer's (1996) reply.

Some time ago we introduced a program of research on judgment under uncertainty, which has come to be known as the *heuristics and biases approach* (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). We suggested that intuitive predictions and judgments are often mediated by a small number of distinctive mental operations, which we called *judgmental heuristics*. For example, a judgment of the prevalence of suicide in a community is likely to be mediated by the ease with which instances come to mind; this is an example of the availability heuristic. And a politician of erect bearing walking briskly to the podium is likely to be seen as strong and decisive; this is an example of judgment by representativeness.

These heuristics, we argued, are often useful but they sometimes lead to characteristic errors or biases, which we and others have studied in some detail. There are several reasons for studying judgmental or perceptual biases. First, they are of interest in their own right. Second, they can have practical implications (e.g., to clinical judgment or intuitive forecasting). Third, the study of systematic error can illuminate the psychological processes that underlie perception and judgment. Indeed, a common method to demonstrate that a particular variable affects a judgment is to establish a correlation between that variable and the judgment, holding the objective criterion constant. For example, the effect of aerial perspective on apparent distance is confirmed by the observation that the same mountain appears

closer on a clear than on a hazy day. Similarly, the role of availability in frequency judgments can be demonstrated by comparing two classes that are equal in objective frequency but differ in the memorability of their instances.

The main goal of this research was to understand the cognitive processes that produce both valid and invalid judgments. However, it soon became apparent that "although errors of judgments are but a method by which some cognitive processes are studied, the method has become a significant part of the message" (Kahneman & Tversky, 1982a, p. 124). The methodological focus on errors and the role of judgmental biases in discussions of human rationality have evoked the criticism that our research portrays the human mind in an overly negative light (see, e.g., Cohen, 1981; Einhorn & Hogarth, 1981; Lopes, 1991). The present article is a response to the latest round in this controversy.

In a series of articles and chapters Gigerenzer (1991, 1993, 1994; Gigerenzer, Hell, & Blank, 1988; Gigerenzer & Murray, 1987, chap. 5) has vigorously attacked the heuristics and biases approach to judgment under uncertainty. Gigerenzer's critique consists of a conceptual argument against our use of the term "bias," and an empirical claim about the "disappearance" of the patterns of judgment that we had documented.

The conceptual argument against the notion of judgmental bias is that there is a disagreement among statisticians and philosophers about the interpretation of probability. Proponents of the Bayesian school interpret probability as a subjective measure of belief. They allow the assignment of probabilities to unique events (e.g., the result of the next Super Bowl, or the outcome of a single toss of a coin) and require these assignments to obey the probability axioms. Frequentists, on the other hand, interpret probability as long-run relative frequency and refuse to assign probability to unique events. Gigerenzer argues that because the concept of subjective probability is controversial in statistics, there is no normative basis for diagnosing such judgments as wrong or biased. Consequently, "biases are not biases" (1991, p. 86), and "heuristics are meant to explain what does not exist" (1991, p. 102).

On the empirical side, Gigerenzer argues that "allegedly sta-

Daniel Kahneman, Department of Psychology and Woodrow Wilson School of Public and International Affairs, Princeton University; Amos Tversky, Department of Psychology, Stanford University.

Amos Tversky died on June 2, 1996.

This work was supported by National Science Foundation Grants SBR-9496347 and SBR-940684 and by National Institute of Mental Health Grant MH53046.

Correspondence concerning this article should be addressed to Daniel Kahneman, Woodrow Wilson School of Public and International Affairs, Prospect Street, Princeton University, Princeton, New Jersey 08544-1013. Electronic mail may be sent via Internet to Kahneman@pucc.princeton.edu.

ble" errors of judgments can be "made to disappear" by two simple manipulations: asking questions in terms of frequencies rather than in terms of probabilities and emphasizing the role of random sampling. He illustrates these claims by a critical discussion of three judgmental biases: base-rate neglect, conjunction errors, and overconfidence. He suggests that the same methods can be used to make other cognitive illusions disappear (p. 300). Gigerenzer concludes that the heuristics and biases approach is a "conceptual dead end" that "has not given us much purchase in understanding judgment under uncertainty" (1991, p. 103).

This article examines the validity of Gigerenzer's critique of heuristics and biases research, which has focused primarily on our work. We make no attempt here to evaluate the achievements and the limitations of several decades of research on heuristics and biases, by ourselves and by others. The next section assesses the accuracy of Gigerenzer's presentation. The following three sections address, in turn, the three phenomena targeted in his critique. The final section provides a summary and discusses the relation between degree of belief and assessments of frequency.

Scope and Accuracy

It is not uncommon in academic debates that a critic's description of the opponent's ideas and findings involves some loss of fidelity. This is a fact of life that targets of criticism should learn to expect, even if they do not enjoy it. In some exceptional cases, however, the fidelity of the presentation is so low that readers may be misled about the real issues under discussion. In our view, Gigerenzer's critique of the heuristics and biases program is one of these cases. The main goal of the present reply is to correct his misleading description of our work and his tendentious presentation of the evidence. The correction is needed to distinguish genuine disagreements from objections to positions we do not hold. In this section we identify some of the major misrepresentations in Gigerenzer's critique.

The scope of the research program is a case in point. The reader of Gigerenzer's critique is invited to believe that the heuristics and biases approach was exclusively concerned with biases in assessments of subjective probability,¹ to which Gigerenzer has had a philosophical objection. However, much of our research has been concerned with tasks to which his objection does not apply. Our 1974 (Tversky & Kahneman) *Science* article, for example, discussed twelve biases. Only two (insensitivity to prior probability of outcomes and overconfidence in subjective probability distributions) involve subjective probability; the other ten biases do not. These include the effect of arbitrary anchors on estimates of quantities, availability biases in judgment of frequency, illusory correlation, nonregressive prediction, and misconceptions of randomness. These findings are not mentioned in Gigerenzer's account of heuristics and biases. Inexplicably, he dismisses the entire body of research because of a debatable philosophical objection to two of twelve phenomena.

The failure to address most of our research has allowed Gigerenzer to offer an erroneous characterization of our normative position as "narrowly Bayesian." Contrary to this description, the normative standards to which we have compared intuitive

judgments have been eclectic and often objective. Thus, we showed that judgments of frequency and estimates of numerical quantities deviate systematically from measured objective values, that estimates of sampling outcomes depart from the values obtained by elementary combinatorial analysis and sampling theory, and that intuitive numerical predictions violate the principle of regression.

Perhaps the most serious misrepresentation of our position concerns the characterization of judgmental heuristics as "independent of context and content" (Gigerenzer et al., 1988) and insensitive to problem representation (Gigerenzer, 1993). Gigerenzer also charges that our research "has consistently neglected Feynman's (1967) insight that mathematically equivalent information formats need not be psychologically equivalent" (Gigerenzer & Hoffrage, 1995, p. 697). Nothing could be further from the truth: The recognition that different framings of the same problem of decision or judgment can give rise to different mental processes has been a hallmark of our approach in both domains.

The peculiar notion of heuristics as insensitive to problem representation was presumably introduced by Gigerenzer because it could be discredited, for example, by demonstrations that some problems are difficult in one representation (probability), but easier in another (frequency). However, the assumption that heuristics are independent of content, task, and representation is alien to our position, as is the idea that different representations of a problem will be approached in the same way. In discussing this point we wrote,

Many adults do not have generally valid intuitions corresponding to the law of large numbers, the role of base rates in Bayesian inference, or the principle of regressive prediction. But it is simply not the case that every problem to which these rules are relevant will be answered incorrectly, or that the rules cannot appear compelling in particular contexts. The properties that make formally equivalent problems easy or hard to solve appear to be related to the mental models, or schemas, that the problems evoke (Kahneman & Tversky, 1982a, pp. 129-130).

We believe that Gigerenzer agrees with our position, and we wonder why it is misrepresented in his writings.

Although we were not able to offer a comprehensive treatment of the process by which different representations and different tasks evoke different heuristics, we investigated this question in several studies. For example, we showed that graphic and verbal representations of a binomial process yield qualitatively different patterns in judgments of frequency (Tversky and Kahneman, 1973), we argued that the use of base-rate data is enhanced when a problem is framed as repetitive rather than unique (Kahneman and Tversky, 1979), and we observed that the impact of base-rate data is increased when these data are given a causal interpretation (Tversky & Kahneman, 1980; see also Ajzen, 1977). We also demonstrated that a representation in terms of absolute frequencies largely eliminated conjunction errors (Tversky & Kahneman, 1983)—a finding that Gigerenzer appears to have appropriated.

¹ For the purposes of the present discussion, we use "subjective probabilities" to refer to probability judgments about unique events.

The major empirical claim in Gigerenzer's critique, that cognitive illusions "disappear" when people assess frequencies rather than subjective probabilities, also rests on a surprisingly selective reading of the evidence. Most of our early work on availability biases was concerned with judgments of frequency (Tversky & Kahneman, 1973), and we illustrated anchoring by inducing errors in judgments of the frequency of African nations in the United Nations (Tversky & Kahneman, 1974). Systematic biases in judgments of frequency have been observed in numerous other studies (e.g., Slovic, Fischhoff, & Lichtenstein, 1982).

These examples should suffice to demonstrate why, in our view, Gigerenzer's reports on our work and on the evidence cannot be taken at face value. Further examples can be found by comparing Gigerenzer's writings (e.g., 1991, 1993, 1994) with our own (in particular, Kahneman & Tversky, 1982a, 1982b; Tversky & Kahneman, 1974, 1983). The position described by Gigerenzer is indeed easy to refute, but it bears little resemblance to ours. It is useful to remember that the refutation of a caricature can be no more than a caricature of refutation.

In the next sections we discuss the three phenomena that Gigerenzer used to illustrate the disappearance of cognitive illusions. In each case we briefly review the original work, then examine his critique in light of the experimental evidence.

Base-Rate Neglect

Intuitive predictions and judgments of probability, we proposed, are often based on the relation of similarity or representativeness between the evidence and possible outcomes. This concept was characterized as follows:

Representativeness is an assessment of the degree of correspondence between a sample and a population, an instance and a category, an act and an actor, or more generally between an outcome and a model. The model may refer to a person, a coin, or the world economy, and the respective outcomes could be marital status, a sequence of heads and tails, or the current price of gold. Representativeness can be investigated empirically by asking people, for example, which of two sequences of heads and tails is more representative of a fair coin or which of two professions is more representative of a given personality (Tversky & Kahneman, 1983, pp. 295–296).

The relation of correspondence or similarity between events, we reasoned, is largely independent of their frequency. Consequently, the base rates of outcomes are likely to have little impact on predictions that are based primarily on similarity or representativeness. We have used the term *base-rate neglect* to describe situations in which a base rate that is known to the subject, at least approximately, is ignored or significantly underweighted. We tested this hypothesis in several experimental paradigms. Gigerenzer's critique of base-rate neglect focuses on a particular design, in which base-rate information is explicitly provided and experimentally manipulated.

In our original experiment, participants read brief descriptions of different individuals, allegedly sampled at random from a group consisting of 30 engineers and 70 lawyers (or 30 lawyers and 70 engineers). Participants assessed the probability that each description referred to an engineer rather than to a lawyer.

The effect of the manipulation of base rate in this experiment was statistically significant, but small. Subsequent studies have identified several factors that enhance the use of base-rate information in this paradigm: presenting the base-rate data after the personality description (Krosnick, Li, & Lehman, 1990), varying base rate across trials (Bar-Hillel & Fischhoff, 1981), and encouraging participants to think as statisticians (Schwarz, Strack, Hilton, & Naderer, 1991). In the same vein, Gigerenzer, Hell, and Blank (1988) reported that repeated random sampling of descriptions increased the use of base rates. The impact of base-rate data was larger in these experiments than in our original study, but less than expected according to Bayes' rule. A fair summary of the evidence is that explicitly stated base rates are generally underweighted but not ignored (see, e.g., Bar-Hillel, 1983).

Gigerenzer, however, reaches a different conclusion. He claims that "If one lets the subjects do the random drawing base-rate neglect disappears" (1991, p. 100). This claim is inconsistent with the data: Underweighting of base-rate was demonstrated in several studies in which participants actually drew random samples from a specified population, such as numbered balls from a bingo cage (Camerer, 1990; Grether, 1980, 1992; Griffin & Dukeshire, 1993). Even in Gigerenzer's own study, all six informative descriptions deviated from the Bayesian solution in the direction predicted by representativeness; the deviations ranged from 6.6% to 15.5% (see Gigerenzer et al., 1988, Table 1, p. 516). Griffin & Dukeshire (1993) observed substantially larger deviations in the same design. To paraphrase Mark Twain, it appears that Gigerenzer's announcement about the disappearance of base-rate neglect is premature.

Gigerenzer notes that "In many natural environments . . . frequencies must be sequentially learned through experience" (1994, p. 149) and suggests that this process allows people to adopt a more effective algorithm for assessing posterior probability. He offers a hypothetical example in which a physician in a nonliterate society learns quickly and accurately the posterior probability of a disease given the presence or absence of a symptom. Indeed, there is evidence that people and other animals often register environmental frequencies with impressive accuracy. However, Gigerenzer's speculation about what a nonliterate physician might learn from experience is not supported by existing evidence. Subjects in an experiment reported by Gluck and Bower (1988) learned to diagnose whether a patient has a rare (25%) or a common (75%) disease. For 250 trials the subjects guessed the patient's disease on the basis of a pattern of four binary symptoms, with immediate feedback. Following this learning phase, the subjects estimated the relative frequency of the rare disease, given each of the four symptoms separately.

If the mind is "a frequency monitoring device," as argued by Gigerenzer (1993, p. 300), we should expect subjects to be reasonably accurate in their assessments of the relative frequencies of the diseases, given each symptom. Contrary to this naive frequentist prediction, subjects' judgments of the relative frequency of the two diseases were determined entirely by the diagnosticity of the symptom, with no regard for the base-rate frequencies of the diseases. Although the participants in this experiment encountered the common disease three times more frequently than the rare disease, they estimated the frequency of disease given symptom, as if the two diseases were equally

likely. Additional evidence for base-rate neglect in this paradigm has been reported by Estes, Campbell, Hatsopoulos, and Hurwitz (1989) and by Nosofsky, Kruschke, and McKinley (1992). Contrary to Gigerenzer's unqualified claim, the replacement of subjective probability judgments by estimates of relative frequency and the introduction of sequential random sampling do not provide a panacea against base-rate neglect.

Most of the research on the use or neglect of base-rate information has focused on situations in which that information is explicitly provided or made observable to the subject. However, the most direct evidence for the role of representativeness in prediction comes from a different experimental situation, which we label the outcome-ranking paradigm. In this paradigm, subjects are given case data about a person (e.g., a personality description) and are asked to rank a set of outcomes (e.g., occupations or fields of study) by different criteria. Subjects in one condition rank the outcomes by representativeness: the degree to which the person resembles the stereotype associated with each outcome. Subjects in the second condition rank the same outcomes by the probability that they apply to the person in question. Subjects in a third group are not given case data; they rank the outcomes by their base rate in the population from which the case is said to be drawn.

The results of several experiments showed that the rankings of outcomes by representativeness and by probability were nearly identical (Kahneman, & Tversky, 1973; Tversky, & Kahneman, 1982). The probability ranking of outcomes did not regress toward the base-rate ranking, even when the subjects were told that the predictive validity of the personality descriptions was low. However, when subjects were asked to make predictions about an individual for whom no personality sketch was given, the probability ranking was highly correlated with the base-rate ranking. Subjects evidently consulted their knowledge of base rates in the absence of case data, but not when a personality description was provided (Kahneman & Tversky, 1973).

Gigerenzer's discussion of representativeness and base-rate neglect has largely ignored the findings obtained in the outcome-ranking paradigm. He dismisses the results of one study involving a particular case (Tom W.) on the grounds that our subjects were not given reason to believe that the target vignette had been randomly sampled (Gigerenzer, 1991, p. 96). Unaccountably, he fails to mention that identical results were obtained in a more extensive study, reported in the same article, in which the instructions explicitly referred to random sampling (Kahneman & Tversky, 1973, Table 2, p. 240).

The outcome-ranking paradigm is especially relevant to Gigerenzer's complaint that we have not provided formal definitions of representativeness or availability and that these heuristics are "largely undefined concepts and can post hoc be used to explain almost everything" (1991, p. 102). This objection misses the point that representativeness (like similarity) can be assessed experimentally; hence it need not be defined a priori. Testing the hypothesis that probability judgments are mediated by representativeness does not require a theoretical model of either concept. The heuristic analysis only assumes that the latter is used to assess the former and not vice versa. In the outcome-ranking paradigm, representativeness is defined operationally by the subjects' ranking, which is compared to an inde-

pendent ranking of the same outcomes by their probability. These rankings of the outcomes rely, of course, on subjects' understanding of the terms probability, similarity, or representativeness. This is a general characteristic of research in perception and judgment: Studies of loudness, fairness, or confidence all rest on the meaning that subjects attach to these attributes, not on the experimenter's theoretical model.

What does all this say about the base-rate controversy and about prediction by representativeness? First, it is evident that subjects sometimes use explicitly mentioned base-rate information to a much greater extent than they did in our original engineer-lawyer study, though generally less than required by Bayes' rule. Second, the use of repeated random sampling is not sufficient to eliminate base-rate neglect, contrary to Gigerenzer's claim. Finally, the most direct evidence for the role of representativeness in intuitive prediction, obtained in the outcome-ranking paradigm, has not been challenged.

Conjunction Errors

Perhaps the simplest and most fundamental principle of probability is the inclusion rule: If A includes B then the probability of B cannot exceed the probability of A; that is, $A \supset B$ implies $P(A) \geq P(B)$. This principle can also be expressed by the conjunction rule, $P(A \& B) \leq P(A)$, since $A \& B$ is a subset of A. Because representativeness and availability are not constrained by this rule, violations are expected in situations where a conjunction is more representative or more available than one of its components. An extensive series of studies (Tversky & Kahneman, 1983) demonstrated such violations of the conjunction rule in both probability and frequency judgments.

The Normative Issue

Imagine a young woman, named Linda, who resembles a feminist, but not a bank teller. You are asked to consider which of two hypotheses is more likely: (a) Linda is a bank teller or (b) Linda is a bank teller who is active in the feminist movement. Gigerenzer insists that there is nothing wrong with the statement that (b) is more probable than (a). He defends this view on the ground that for a frequentist this proposition is meaningless and argues that "it would be foolish to label these judgments 'fallacies'" (1991, p. 95). The refusal to apply the concept of probability to unique events is a philosophical position that has some following among statisticians, but it is not generally shared by the public. Some weather forecasters, for instance, make probabilistic predictions (e.g., there is 50% chance of rain on Sunday), and the sports pages commonly discuss the chances of competitors in a variety of unique contests. Although lay people are often reluctant to express their degree of belief by a number, they readily make comparative statements (e.g., Brown is more likely than Green to win the party's nomination), which refer to unique events and are therefore meaningless to a radical frequentist.

Although Gigerenzer invokes the meaningfulness argument with great conviction, his position on the issue is problematic. On the one hand, he surely does not regard statements of subjective probability as meaningless; he has even collected such judgments from subjects. On the other hand, he invokes the

argument that subjective probabilities are meaningless to deny that these judgments are subject to any normative standards. This position, which may be described as normative agnosticism, is unreasonably permissive. Is it not a mistake for a speaker to assign probabilities of .99 both to an event and to its complement? We think that such judgments should be treated as mistaken; they violate accepted constraints on the use of probability statements in everyday discourse.

Normative agnosticism is particularly inappropriate in the case of the conjunction rule. First, the application of this rule does not require numerical estimates, only an ordinal judgment of which of two events is more probable. Second, the normative basis for the conjunction rule is essentially logical: If the conjunction $A \& B$ is true then A must also be true, but the converse does not hold. In support of his agnostic position, Gigerenzer cites von Mises's (1928/1957) statement that

We can say nothing about the probability of death of an individual even if we know his condition of life and health in detail. The phrase "probability of death," when it refers to a single person, has no meaning at all for us (p. 11).

Whether or not it is meaningful to assign a definite numerical value to the probability of survival of a specific individual, we submit (a) that this individual is less likely to die within a week than to die within a year and (b) that most people regard the preceding statement as true—not as meaningless—and treat its negation as an error or a fallacy.

Normative agnosticism is even harder to justify when violations of the conjunction rule lead to a preference for a dominated course of action. Several such cases have been documented. For example, we found that most subjects chose to bet on the proposition that Linda is a feminist bank teller rather than on the proposition that she is a bank teller. We also found that most subjects violated the conjunction rule in betting on the outcomes of a dice game involving real payoffs (Tversky & Kahneman, 1983). Further evidence for conjunction errors in choice between bets has been presented by Bar-Hillel and Neter (1993) and by Johnson, Hershey, Meszaros, and Kunreuther (1993). Would Gigerenzer's agnosticism extend to the choice of a dominated option? Or would he agree that there are, after all, some biases that need to be explained?

The Descriptive Issue

Gigerenzer's major empirical claim is that violations of the conjunction rule are confined to subjective probabilities and that they do not arise in judgments of frequencies. This claim is puzzling because the first demonstration in our conjunction paper involves judgments of frequency. Subjects were asked to estimate the number of "seven-letter words of the form '-----n-' in 4 pages of text." Later in the same questionnaire, these subjects estimated the number of "seven-letter words of the form '-- -ing' in 4 pages of text." Because it is easier to think of words ending with "ing" than to think of words with "n" in the next-to-last position, availability suggests that the former will be judged more numerous than the latter, in violation of the conjunction rule. Indeed, the median estimate for words ending with "ing" was nearly three times higher than for words with "n" in the next-to-the-last position. This finding is a counter-

example to Gigerenzer's often repeated claim that conjunction errors disappear in judgments of frequency, but we have found no mention of it in his writings.

Early in our investigation of the conjunction problem, we believed that violations of the conjunction rule only occur when the critical events are evaluated independently, either by different subjects or by the same subject on different occasions. We expected that subjects would conform to the inclusion rule when asked to judge the probability or frequency of a set and of one of its subsets in immediate succession. To our surprise, violations of the conjunction rule turned out to be common even in this case; the detection of inclusion and the appreciation of its significance were evidently more difficult than we had thought.

We therefore turned to the study of cues that may encourage extensional reasoning and developed the hypothesis that the detection of inclusion could be facilitated by asking subjects to estimate frequencies. To test this hypothesis, we described a health survey of 100 adult men and asked subjects, "How many of the 100 participants have had one or more heart attacks?" and "How many of the 100 participants both are over 55 years old and have had one or more heart attacks?" The incidence of conjunction errors in this problem was only 25%, compared to 65% when the subjects were asked to estimate percentages rather than frequencies. Reversing the order of the questions further reduced the incidence to 11%. We reasoned that the frequency formulation may lend itself to a spatial representation, in terms of tokens or areas, which makes the relation of set inclusion particularly salient. This representation seems less natural for percentages, which require normalization.²

Gigerenzer has essentially ignored our discovery of the effect of frequency and our analysis of extensional cues. As primary evidence for the "disappearance" of the conjunction fallacy in judgments of frequency, he prefers to cite a subsequent study by Fiedler (1988), who replicated both our procedure and our findings, using the bank-teller problem. There were relatively few conjunction errors when subjects estimated in immediate succession the number of bank tellers and of feminist bank tellers, among 100 women who fit Linda's description. Gigerenzer concludes that "the conceptual distinction between single events and frequency representations is sufficiently powerful to make this allegedly-stable cognitive illusion disappear" (1993, p. 294). In view of our prior experimental results and theoretical discussion, we wonder who alleged that the conjunction fallacy is stable under this particular manipulation.

It is in the nature of both visual and cognitive illusions that there are conditions under which the correct answer is made transparent. The Müller-Lyer Illusion, for example, "disappears" when the two figures are embedded in a rectangular frame, but this observation does not make the illusion less interesting. The hypothesis that people use a heuristic to answer a

² Cosmides and Tooby (1996) have shown that a frequentistic formulation also helps subjects solve a base-rate problem that is quite difficult when framed in terms of percentages or probabilities. Their result is readily explained in terms of extensional cues to set inclusion. These authors, however, prefer the speculative interpretation that evolution has favored reasoning with frequencies but not with percentages.

difficult question does not entail that they are blind to a salient cue that makes the correct answer obvious. We have argued that the frequency formulation provides a powerful cue to the relation of inclusion between sets that are explicitly compared or evaluated in immediate succession. This extensional cue is not available to participants who evaluate the sets separately in a between-subjects design. We predict, therefore, that the frequency formulation, which greatly reduces the incidence of conjunction errors in a direct comparison, will not have much effect in a between-subjects design. If, on the other hand, violations of the conjunction rule in the bank-teller problem are confined to judgments about single events (as suggested by Gigerenzer) frequency judgments should obey the rule even in a between-subjects design. To test the opposing predictions, we presented subjects with the following problem.

Linda is in her early thirties. She is single, outspoken, and very bright. As a student she majored in philosophy and was deeply concerned with issues of discrimination and social justice.

Suppose there are 1,000 women who fit this description. How many of them are

- (a) high school teachers?
- (b) bank tellers? or
- (c) bank tellers and active feminists?"

One group of Stanford students ($N = 36$) answered the above three questions. A second group ($N = 33$) answered only questions (a) and (b), and a third group ($N = 31$) answered only questions (a) and (c). Subjects were provided with a response scale consisting of 11 categories in approximately logarithmic spacing. As expected, a majority (64%) of the subjects who had the opportunity to compare (b) and (c) satisfied the conjunction rule. In the between-subjects comparison, however, the estimates for feminist bank tellers (median category: "more than 50") were significantly higher than the estimates for bank tellers (median category: "13-20," $p < .01$ by a Mann-Whitney test). Contrary to Gigerenzer's position, the results demonstrate a violation of the conjunction rule in a frequency formulation. These findings are consistent with the hypothesis that subjects use representativeness to estimate outcome frequencies and edit their responses to obey class inclusion in the presence of strong extensional cues. The finding that the conjunction rule is applied in direct comparisons, but not in between-subjects experiments, indicates that the key variable that controls adherence to the conjunction rule is not the contrast between single events and frequencies, but the opportunity to detect the relation of set inclusion.

The Methodological Issue

The preceding study illustrates the important difference between within-subject and between-subjects designs, which is sometimes overlooked in discussions of judgment and of judgment errors. The within-subject design, in which critical items are presented in immediate succession, provides subjects with information that is not available to subjects in a between-subjects design.³ First, it often reveals the intent of the researcher, by drawing attention to the independent variable that is manipulated. Second, the subject has a chance to detect and correct errors and inconsistencies in the responses to different items.

The two designs address different questions, especially in cases of conflict between a judgmental heuristic (e.g., representativeness) and a compelling formal principle (e.g., the conjunction rule). A between-subjects design provides a clean test of the hypothesis that subjects rely on a given heuristic. The within-subjects design addresses the question of how the conflict between the heuristic and the rule is resolved. In the case of the conjunction rule, the evidence shows that sometimes the heuristic prevails, sometimes the rule, depending on the sophistication of the subjects, the transparency of the problem, or the effectiveness of the extensional cues (Tversky & Kahneman, 1983). Thus, the between-subjects design (indirect test) is appropriate when we wish to understand "pure" heuristic reasoning; the within-subject design (direct test) is appropriate when we wish to understand how conflicts between rules and heuristics are resolved.⁴

The direct and the indirect tests have a somewhat different normative status. Suppose two different groups of subjects are randomly assigned to assess the probability (or frequency) of an event A or of a conjunction A & B. In this case, it is not possible to say that any particular subject committed a conjunction error, even if all judgments of A & B are higher than all judgments of A. Nevertheless, such a finding (or even a less extreme result, as obtained in the above experiment) establishes that subjects have a disposition to answer the two questions inconsistently; they do not derive their answers from a coherent structure of estimates or beliefs. Gigerenzer appears to deny the relevance of the between-subjects design on the ground that no individual subject can be said to have committed an error. In our view, this is hardly more reasonable than the claim that a randomized between-subject design cannot demonstrate that one drug is more effective than another because no individual subject has experienced the effects of both drugs.

Overconfidence

In the calibration paradigm, subjects answer multiple-choice questions and state their probability, or confidence, that they have selected the correct answer to each question. The subjects in these experiments are normally instructed to use the probability scale so that their stated confidence will match their expected accuracy. Nevertheless, these studies often report that confidence exceeds accuracy. For example, when subjects express 90% confidence, they may be correct only about 75% of the time (for reviews, see Keren, 1991; Lichtenstein, Fischhoff, & Phillips, 1982; Yates, 1990). Overconfidence is prevalent but not universal: It is generally eliminated and even reversed for very easy items. This phenomenon, called the difficulty effect, is an expected consequence of the definition of overconfidence as the difference between mean confidence and overall accuracy.

Consistent with his agnostic normative stance, Gigerenzer ar-

³ The two designs also induce different conversational implicatures. For a discussion of this issue, including several demonstrations that violations of the conjunctions rule cannot be attributed to linguistic ambiguities, see Tversky and Kahneman (1983).

⁴ It is sometimes possible to conduct an indirect test in a within-subject design by separating the critical items spatially or temporally so as to avoid a direct comparison.

gues that overconfidence should not be viewed as a bias because judgments of confidence are meaningless to a frequentist. This argument overlooks the fact that in most experiments the subjects were explicitly instructed to match their stated confidence to their expected accuracy. The presence of overconfidence therefore indicates that the subjects committed at least one of the following errors: (a) overly optimistic expectation or (b) a failure to use the scale as instructed. Proper use of the probability scale is important because this scale is commonly used for communication. A patient who is informed by his surgeon that she is 99% confident in his complete recovery may be justifiably upset to learn that when the surgeon expresses that level of confidence, she is actually correct only 75% of the time. Furthermore, we suggest that both surgeon and patient are likely to agree that such a calibration failure is undesirable, rather than dismiss the discrepancy between confidence and accuracy on the ground that "to compare the two means comparing apples and oranges" (Gigerenzer, 1991, p. 88).

Gigerenzer's descriptive argument consists of two points. First, he attributes overconfidence to a biased selection of items from a domain and predicts that overconfidence will vanish when items are randomly selected from a natural reference class. Second, he argues that overconfidence disappears when people assess relative frequency rather than subjective probability. We discuss these points in turn.

Gigerenzer writes,

If the general knowledge questions were a representative sample from the knowledge domain, zero overconfidence would be expected. . . . However, general knowledge questions typically are not representative samples from some domain of knowledge, but are selected to be difficult or even misleading. . . . "overconfidence bias" results as a consequence of selection, not of some deficient mental heuristics (1993, p. 304).

This account of overconfidence, which draws on the theory of probabilistic mental models (Gigerenzer, Hoffrage, & Kleinbolting, 1991), encounters both conceptual and empirical difficulties. First, it is far from clear in most cases what constitutes a random or a representative set of questions for a given knowledge domain (e.g., geography, politics, or baseball) and how to construct such a sample. Second, although the deliberate selection of difficult or surprising items can produce spurious overconfidence, it is not generally the case that overconfidence is eliminated by random sampling of items.

Several researchers have selected at random pairs of items (e.g., cities) from some reference class and asked subjects to indicate, say, which city has the larger population and to express their confidence in each answer. Some experiments, in which the questions were relatively easy, indicated no overconfidence (Gigerenzer et al., 1991; Juslin, 1994); but substantial overconfidence was observed in other studies, in which the questions were slightly harder (Griffin & Tversky, 1992; Liberman, 1996). Not surprisingly, perhaps, difficulty emerges as the major, albeit not the sole, determinant of overconfidence, even when the items were selected at random. Contrary to Gigerenzer's prediction, random sampling of items is not sufficient to eliminate overconfidence. Additional support for this conclusion comes from observations of overconfidence in the prediction of natural events (e.g., economic recessions, medical diag-

noses, bridge tournaments), where biased selection of items is not an issue. For further discussion, see Brenner, Koehler, Liberman, and Tversky (in press); Keren and Van Bolhuis (1996).

Let us turn now to the relation between confidence judgments and frequency estimates. May (1987, 1988) was the first to report that whereas average confidence for single items generally exceeds the percentage of correct responses, people's estimates of the percentage (or frequency) of items that they have answered correctly is generally lower than the actual number. In her study of students' knowledge of psychology, the overall percentage of correct predictions was 72%, mean confidence was 81%, and the mean estimate of the percentage of correct responses was only 63%. These data yield 9% overconfidence in judgments of single items and 9% underconfidence in the estimation of the percentage of correct answers. Subsequent studies (e.g., Gigerenzer et al., 1991; Griffin & Tversky, 1992; Sniezek, Paese, & Switzer, 1990) have reported a similar pattern although the degree of underconfidence varied substantially across domains.

Gigerenzer portrays the discrepancy between individual and aggregate assessments as incompatible with our theoretical position, but he is wrong. On the contrary, we drew a distinction between two modes of judgment under uncertainty, which we labeled the inside and the outside views (Kahneman & Tversky, 1979, 1982b; Kahneman & Lovallo, 1993). In the outside view (or frequentistic approach) the case at hand is treated as an instance of a broader class of similar cases, for which the frequencies of outcomes are known or can be estimated. In the inside view (or single-case approach) predictions are based on specific scenarios and impressions of the particular case. We proposed that people tend to favor the inside view and as a result underweight relevant statistical data. For example, students (as well as professors) commonly underestimate the amount of time they need to complete academic projects although they are generally aware of their susceptibility to an optimistic bias (Buehler, Griffin, & Ross, 1994). The inside and outside views bring different evidence to mind. As Griffin and Tversky (1992) put it

A judgment of confidence in a particular case, we propose, depends primarily on the balance of arguments for and against a specific hypothesis. Estimated frequency of correct prediction, on the other hand, is likely to be based on a general evaluation of the difficulty of the task, the knowledge of the judge, or past experience with similar problems (p. 431).

Because people tend to adopt the inside view, they can maintain a high degree of confidence in the validity of specific answers even when they know that their overall hit rate is low. We first observed this phenomenon in the context of the prediction of success in officer training school (Kahneman & Tversky, 1973), and we called it the illusion of validity.

The preceding discussion should make it clear that, contrary to Gigerenzer's repeated claims, we have neither ignored nor blurred the distinction between judgments of single and of repeated events. We proposed long ago that the two tasks induce different perspectives, which are likely to yield different estimates, and different levels of accuracy (Kahneman and Tversky, 1979). As far as we can see, Gigerenzer's position on this issue is not different from ours, although his writings create the

opposite impression. Our disagreement here is normative, not descriptive. We believe that subjective probability judgments should be calibrated, whereas Gigerenzer appears unwilling to apply normative criteria to such judgments.

Discussion

As this review has shown, Gigerenzer's critique employs a highly unusual strategy. First, it attributes to us assumptions that we never made (e.g., that judgmental heuristics are independent of content and context or that judgments of probability and of frequency yield identical results). Then it attempts to refute our alleged position by data that either replicate our prior work (extensional cues reducing conjunction errors) or confirm our theoretical expectations (discrepancy between individual and global measures of overconfidence). These findings are presented as devastating arguments against a position that, of course, we did not hold. Evidence that contradicts Gigerenzer's conclusion (base-rate neglect with explicit random sampling; conjunction errors in frequency judgments) is not acknowledged and discussed, as is customary; it is simply ignored. Although some polemic license is expected, there is a striking mismatch between the rhetoric and the record in this case.

Gigerenzer's polemics obscure a surprising fact: There is less psychological substance to his disagreement with our position than meets the eye. Aside from the terminological question of whether terms such as "error" or "bias" can be applied to statements of subjective probability, the major empirical point made by Gigerenzer is that the use of frequency reliably makes cognitive illusions "disappear." Taken at face value, this statement is just wrong. Because Gigerenzer must be aware of the evidence that judgments of frequency and judgments based on frequency are subject to systematic error, a charitable interpretation of his position is that he has overstated his case by omitting relevant quantifiers. Thus, some cognitive illusions (not all) are sometimes reduced (not made to disappear) in judgments of frequency. This position is much more faithful to the evidence; it is also no longer in conflict with what we have said on this topic, which may be summarized as follows: (a) The adoption of an "outside view" that brings to bear the statistics of past cases can sometimes improve the accuracy of judgment concerning a single case (Kahneman & Lovalló, 1993; Kahneman & Tversky, 1979). (b) The frequency formulation sometimes makes available strong extensional cues that subjects can use to avoid conjunction errors in a within-subject design. (c) There are substantial biases in judgments of frequency, often the same biases that affect judgments of probability (Tversky & Kahneman, 1983).

After the record is corrected, some differences of opinion and emphasis on matters of psychological substance remain. Gigerenzer emphasizes both the accuracy and the significance of judgments of frequency and underplays the importance of subjective probability; he also believes that subjective probabilities can be explained in terms of learned frequencies. We do not share these views. A long history of research shows that judgment errors often persist in situations that provide ready access to relevant frequency data. L. J. Chapman and J. P. Chapman's (1967, 1969) studies of illusory correlation offer a compelling experimental demonstration of the persistence of errors in-

duced by representativeness. Lay subjects and clinical psychologists who were shown data about a series of individual cases perceived illusory correlations between clinical diagnoses and representative symptoms (e.g., paranoia and peculiar eyes in the Draw-A-Person test). Illusory correlation was resistant to contradictory data, and it prevented the judges from detecting relationships that were in fact present. Similarly, studies of the belief in the hot hand in basketball (Gilovich, Vallone, & Tversky, 1985; Tversky & Gilovich, 1989) have shown that people "see" a positive serial correlation between the outcomes of successive shots, even when no such correlation is present in the data. These findings do not imply that people are incapable of learning the correct contingencies; they only show, contrary to a naive frequentist position, that some significant judgmental biases are not readily corrected by the observation of natural frequencies.

Subjective judgments of probability are important because action is often based on beliefs regarding single events. The decisions of whether or not to buy a particular stock, undergo a medical operation, or go to court depend on the degree to which the decision maker believes that the stock will go up, the operation will be successful, or the court will decide in her favor. Such events cannot be generally treated as a random sample from some reference population, and their judged probability cannot be reduced to a frequency count. Studies of frequency estimates are unlikely to illuminate the processes that underlie such judgments. The view that "both single-case and frequency judgments are explained by learned frequencies (probability cues), albeit by frequencies that relate to different reference classes" (Gigerenzer, 1991, p. 106) appears far too restrictive for a general treatment of judgment under uncertainty. First, this treatment does not apply to events that are unique for the individual and therefore excludes some of the most important evidential and decision problems in people's lives. Second, it ignores the role of similarity, analogy, association, and causality. There is far more to inductive reasoning and judgment under uncertainty than the retrieval of learned frequencies.

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. (1983). The base rate fallacy controversy. In R. W. Scholz (Ed.), *Decision making under uncertainty* (pp. 39-61). Amsterdam: North-Holland.
- Bar-Hillel, M., & Fischhoff, B. (1981). When do base rates affect predictions? *Journal of Personality and Social Psychology*, 41, 671-680.
- Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in stereotype judgments. *Journal of Personality and Social Psychology*, 65, 1119-1131.
- Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (in press). Overconfidence in probability and frequency judgments. *Organizational Behavior and Human Decision Process*.
- Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67, 366-381.
- Camerer, C. (1990). Do markets correct biases in probability judgment? Evidence from market experiments. In L. Green & J. H. Kagel (Eds.), *Advances in behavioral economics* (Vol. 2, pp. 126-172). Greenwich, CT: JAI Press.

- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 73*, 193–204.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic steps. *Journal of Abnormal Psychology, 74*, 271–280.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences, 4*, 317–331.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1–73.
- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32*, 53–88.
- Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 556–571.
- Feynman, R. (1967). *The character of physical law*. Cambridge, MA: MIT Press.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50*, 123–129.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology*, (Vol. 2, pp. 83–115). Chichester, England: Wiley.
- Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 284–313). London: Routledge.
- Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–161). New York: Wiley.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky (1996). *Psychological Review, 103*, 592–596.
- Gigerenzer, G., Hoffrage, U., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 513–525.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.
- Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gilovich, T., Vallone, B., & Tversky, A. (1985). The hot hand in basketball: On the misconception of random sequences. *Cognitive Psychology, 17*, 295–314.
- Gluck, M., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227–247.
- Grether, D. M. (1980). Bayes' rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics, 95*, 537–557.
- Grether, D. M. (1992). Testing Bayes' rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior and Organization, 17*, 31–57.
- Griffin, D., & Dukeshire, S. (1993). *The role of visual random sampling in base rate use and neglect: A methodological critique*. Unpublished manuscript, University of Waterloo, Waterloo, Ontario, Canada.
- Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.
- Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty, 7*, 35–51.
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226–246.
- Kahneman, D., & Lovallo, D. (1993). Bold forecasting and timid decisions: A cognitive perspective on risk taking. *Management Science, 39*, 17–31.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.
- Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science, 12*, 313–327.
- Kahneman, D., & Tversky, A. (1982a). On the study of statistical intuitions. *Cognition, 11*, 123–141.
- Kahneman, D., & Tversky, A. (1982b). Variants of uncertainty. *Cognition, 11*, 143–157.
- Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*(3), 217–273.
- Keren, G., & Van Bolhuis, J. (1996). *On the ecological validity of calibration studies*. Unpublished manuscript, University of Technology, Eindhoven, The Netherlands.
- Krosnick, J. A., Li, F., & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology, 59*, 1140–1152.
- Liberman, V. (1996). *Local and global judgments of confidence*. Unpublished manuscript, Open University, Tel Aviv, Israel.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, England: Cambridge University Press.
- Lopes, L. (1991). The rhetoric of irrationality. *Theory and Psychology, 1*, 65–82.
- May, R. S. (1987). *Calibration of subjective probabilities: A cognitive analysis of inference processes in overconfidence* (in German). Frankfurt, Germany: Peter Lang.
- May, R. S. (1988). Overconfidence in overconfidence. In A. Chikan, J. Kindler, & I. Kiss (Eds.), *Proceedings of the 4th FUR Conference*. Dordrecht, The Netherlands: Kluwer.
- Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 211–233.
- Schwarz, N., Strack, F., Hilton, D. J., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition, 9*(1), 67–84.
- Slovic, P., Fischhoff, G., & Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 463–489). Cambridge, England: Cambridge University Press.
- Sniezek, J. A., Paese, P. W., & Switzer, F. S., III. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46*, 264–282.
- Tversky, A., & Gilovich, T. (1989). The hot hand: Statistical reality or cognitive illusion? *Chance, 2*(1), 16–21.

- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 84-98). Hillsdale, NJ: Erlbaum.
- Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84-98). Cambridge, England: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 91, 293-315.
- Von Mises, R. (1957). *Probability, statistics, and truth*. London: Allen & Unwin. (Original work published 1928)
- Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall.

Received December 30, 1994

Revision received June 27, 1995

Accepted June 29, 1995 ■

Postscript

Gigerenzer's (1996) reply, which follows, reiterates his objections to our work without answering our main arguments. His treatment of the conjunction effect illustrates the formula he uses to dismiss our results: reinterpret a single example (Linda, within-subject); ignore documented cases in which this interpretation fails; discard between-subjects experiments because they allegedly cannot demonstrate error.

This formula will not do. Whether or not violations of the conjunction rule in the between-subjects versions of the Linda and "ing" problems are considered errors, they require explanation. These violations were predicted from representativeness and availability, respectively, and were observed in both frequency and probability judgments. Gigerenzer ignores this evidence for our account and offers no alternative.

Gigerenzer rejects our approach for not fully specifying the condi-

tions under which different heuristics control judgment. Much good psychology would fail this criterion. The Gestalt rules of similarity and good continuation, for example, are valuable although they do not specify grouping for every display. We make a similar claim for judgmental heuristics.

Gigerenzer legislates process models as the primary way to advance psychology. Such legislation is unwise. It is useful to remember that the qualitative principles of Gestalt psychology long outlived premature attempts at modeling. It is also unwise to dismiss 25 years of empirical research, as Gigerenzer does in his conclusion. We believe that progress is more likely to come by building on the notions of representativeness, availability, and anchoring than by denying their reality.