Psychological Bulletin
1990, Vol. 107, No. 3, 296–310

0033-2909/90/$00.75

Why We Still Use Our Heads Instead of Formulas: Toward an Integrative Approach

Benjamin Kleinmuntz
University of Illinois at Chicago

This review begins with a discussion of Meehl's (1957) query regarding when to use one's head (i.e., intuition) instead of the formula (i.e., statistical or mechanical procedure) for clinical prediction. It then describes the controversy that ensued and analyzes the complexity and contemporary relevance of the question itself. Going beyond clinical inference, it identifies select cognitive biases and constraints that cause decision errors, and proposes remedial correctives. Given that the evidence shows cognition to be flawed, the article discusses the linear regression, Bayesian, signal detection, and computer approaches as possible decision aids. Their cost–benefit trade-offs, when used either alone or as complements to one another, are examined and evaluated. The critique concludes with a note of cautious optimism regarding the formula's future role as a decision aid and offers several interim solutions.

More than three decades ago, an article by Meehl (1957) asked, "When shall we use our heads instead of the formula?" He replied that if people have a formula, then they should use their heads only very, *very* seldom. *Heads* in the article's title refers to the processing of data clinically, subjectively, or intuitively; *formula* refers to its nonjudgmental, mathematical, statistical, or mechanical combination.

One purpose of this review is to extend Meehl's (1957) query beyond clinical psychology. Accordingly, it draws examples from medicine, polygraphy, engineering, finance, accounting, management, game playing, and revenue collection. Another objective is to explore the complexity of the question itself. This article addresses six main issues: (a) Cognition is flawed; (b) the flaws are remediable, given proper training and closer correspondence between intuition and task environments; (c) analytical reasoning, formulas, or both can improve thinking; (d) formulas can be used as a standard with which to compare cognition; (e) judgment can be aided when used together with formulas; and (f) there are cost–benefit trade-offs associated with using unaided as well as aided intuition.

Preparation of this review was partially supported by National Science Foundation Grant SES 83-19138 and National Library of Medicine Grant 1RO1 LM04583-02.

Special thanks are due to Don N. Kleinmuntz for his critical comments, but he is not to be held accountable for the biases, errors, and sins of his father; and thanks are also due to Hal Arkes, Caryn Christensen, Tim de Dombal, Ward Edwards, Arthur Elstein, Kenneth Hammond, Robin Hogarth, John Payne, and Gordon Pitz for their help.

One whose comments would have improved this article immeasurably, the late Hillel J. Einhorn, has departed us all too soon on January 8, 1987. His presence, however, is evident throughout. I miss him much, both as an intellectual resource and as a good friend.

Correspondence concerning this article should be addressed to Benjamin Kleinmuntz, Department of Psychology, Box 4348, University of Illinois, Chicago, Illinois 60680.

Perhaps the most important objective of this critique is to raise an old issue that is again of contemporary interest. For, as Einhorn (1988) and Simon (1986) recently observed, increased access to computers and other decision supports invites new comparisons of human with machine intelligence. This being the case, it is timely to assess the pros and cons of using intuition with and without decision aids. Although this article begins on a divisive note—whether to use the head *or* the formula—it concludes by recommending that *both* be used. By so doing, I aspire to foster further research aimed at exploring how, when, and where using the head, the formula, or both will improve decisions.

Before embarking on this inquiry, I simply want to note that in what follows the terms *prediction, forecasting, cognition, thought, inference, judgment, choice, diagnosis, decision making, intuition, reasoning,* and *problem solving* are often used interchangeably, although more careful distinctions have been made elsewhere (e.g., see Anderson & Reder, 1987; Billings & Scherer, 1988; Einhorn & Hogarth, 1982; Tversky, Sattah, & Slovic, 1988). In this article, such distinctions and precision are unnecessary.

Statistical *or* Intuitive Judgment

Argument and Some Evidence for Formulas

One of the earliest calls for the scientific study of judgment came from Meehl's (1954) influential book, *Clinical Versus Statistical Prediction.* He argued that many judgments are best made statistically, not intuitively. He reviewed 20 empirical studies comparing the two prediction modes. Only once was intuition better than statistics. Meehl (1965) later increased this box-score tally to 51 studies, of which 33 favored the head; and 17 demonstrated "approximate equality" of the two approaches (see J. S. Wiggins, 1973, pp. 182–189). Using a somewhat different framework, Sawyer (1966) reviewed 45 studies. He found none in which clinical prediction excelled.

This form of scorekeeping, which became an integral part of the so-called clinical versus statistical prediction controversy, has all but subsided in intensity over the years, albeit with an occasional spontaneous recovery (e.g., Dawes, 1976, 1979, 1988; Einhorn, 1986; Goldberg, in press; Holt, 1978, 1986; B. Kleinmuntz, in press; Meehl, 1986; Sarbin, 1986). An early study in this debate, that by Goldberg (1965), was particularly important for the formula side. It provided strong evidence that in predicting a dichotomous diagnosis, a simple linear composite of five Minnesota Multiphasic Personality Inventory scales outperformed the best from among 13 clinicians (for decision rule alternatives to this formula, see Alexander & Kleinmuntz, 1962; Meehl & Dahlstrom, 1960). Other studies, using various mechanical modes of information processing, obtained similar results (e.g., see Dawes, 1971; Goldberg, 1969, 1971; Grebstein, 1963; B. Kleinmuntz, 1963; Sawyer, 1966). Thus, the evidence seemed clearly to favor the formula's use in personality assessment.

Counterargument but No Evidence for Heads

One of Meehl's challengers over the years has been Holt (1958, 1970, 1978, 1986). He has found Meehl's analysis of judgmental deficiency disconcerting. Holt believed that meaningful person assessment must involve subjectivity. Evidently, the clinician is necessary so as to perceive, integrate, synthesize, and hence intuit a theory of the person being assessed.

The evidence favoring such intuition, however, has been meager. For example, the one study that showed the head to be equally as good as, if not better than, the formula (e.g., Lindzey, 1965; see also Meehl, 1965) was properly criticized for its methodological flaws (e.g., Goldberg, 1968a). Its current status is that, at best, it can be considered a tie (see Dawes, 1976, 1979, 1988; Wiggins, 1973, p. 185). Holt's position, therefore, being devoid of empirical support, is untenable.

In any case, Meehl's (1957, 1967) view does not differ radically from Holt's. He clearly stated that humans excel over formulas in selected predictive tasks. For example, he (1954, p. 24) illustrated this with the case of Professor X, for whom a prediction equation yields a .90 probability of going to the movie on a particular night. If Professor X, however, were to have just broken his leg, then the equation would not hold. The broken leg exemplifies the importance of special cases.

Writing on the topic recently, Meehl (1986) had this to say about intuition: "95% of the ordinary decisions made by working practitioners [in mental health settings] are not comparable in richness and subtlety to that of a good psychoanalytic hour" (p. 373). On behalf of the formula, he observed, "When you check out at a supermarket, you don't eyeball the heap of purchases and say to the clerk, 'Well it looks to me as if it's about $17.00 worth; what do you think?' The clerk adds it up" (p. 372). It seems, then, that Meehl is not the wicked actuary often portrayed by some on the "clinical" side.

Controversy Fallout: Cognition at Center Stage

These polemics aside, however, Meehl's (1954, 1957, 1959, 1960, 1965, 1967, 1986) main contribution over the years has been to place judgment at center stage. He did not deal with it at a general philosophical level, nor did he provide just another tool for the statistical side. Rather, he provided a sound rationale and empirical evidence for the scientific scrutiny of judgment.

This facet of his argument inspired different kinds of studies. Some laboratories, for instance, became less concerned with the relative *accuracy* of the two methods than with what Meehl (1960) called "the cognitive activity of the clinician." This shifted the focus onto the inferential *process*. Hoffman (1960, 1968; see also Wiggins & Hoffman, 1968), for example, proposed linear, configural, and analysis of variance models to describe judgment. Hoffman did not claim that people actually combine information mathematically, nor did he claim that such models outperform humans; instead, he showed the models' descriptive and predictive powers. Perhaps more important, however, Hoffman modeled experts at their best.

Similarly, Hammond (1955) and his associates (e.g., Hammond, Hursch, & Todd, 1964; Hammond & Summers, 1965), somewhat influenced by Meehl but more so by Brunswik's (1952, 1955, 1956) lens model paradigm, also modeled judgment. Hammond and associates showed that the lens model can approximate one's weighting of the cues or predictors in a task. Furthermore, their scheme, with some modification (see Tucker, 1964), demonstrated that the formula captures the optimal or suboptimal use of available environmental cues. It does so by specifying the lower and upper limits of judgmental capability, given the cues of a decision problem. Lens modeling has been explored in a large variety of clinical and nonclinical contexts (e.g., Brady & Rappoport, 1973; Brehmer, 1972; Camerer, 1981; Dudycha & Naylor, 1966; Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Hammond, 1965, 1978; Hammond, Summers, & Deane, 1973; Slovic & Lichtenstein, 1971; Szucko & Kleinmuntz, 1981) and is discussed extensively elsewhere (e.g., Hammond, Stewart, Brehmer, & Steinmann, 1975).

Thus, the study of judgment and cognition, which was once a concern only within clinical psychology now extends beyond that narrow bound. It is evident in such diverse domains as, for instance, medicine (Blois, 1980; Einhorn, 1972, 1974; B. Kleinmuntz & Elstein, 1987), polygraphy (B. Kleinmuntz & Szucko, 1984; Lykken, 1981), security investment (Slovic, 1972), legal adjudication (Hastie, Penrod, & Pennington, 1983; Saks & Kidd, 1980; Wrightsman, 1987), auditing (R. H. Ashton, Kleinmuntz, Sullivan, & Tomassini, 1989), and management (Blattberg & Hoch, in press). More generally, its importance has been acknowledged in the judgment and decision literature at large (e.g., Arkes & Hammond, 1986; Edwards & von Winterfeldt, 1986; Fischhoff, 1987; Hogarth, 1987; Tversky & Kahneman, 1982). The concern about flawed intuition has also laid the groundwork for using heads in combination with formulas (Blattberg & Hoch, in press; Einhorn, 1972; Hogarth, 1978; Sawyer, 1966), as I indicate later.

In sum, then, I have now made five observations about judgment: (a) It is flawed, (b) it can be outperformed, (c) it can be modeled by formulas, (d) it is worthy of further study, and (e) it may be used in combination with formulas. How it is flawed, what can be done about it, and where and how it is useful in and of itself and in combination with formulas are addressed in the remainder of this article.

Unaided Judgment

Some time ago, Simon (1955, 1956) introduced the concept of *bounded rationality.* This is the idea that cognition is limited vis à vis the economist's normative (or rational) model, by which Simon meant that people do not think rationally because doing so, he posited, requires excessive cognitive effort. Instead, they *satisfice;* that is, they set a criterion acceptance level and then use a simplifying decision strategy or heuristic to meet that level. The most important psychological statement and research on this topic thereafter, according to one source (Jungermann, 1986), have come from Tversky and Kahneman (e.g., 1974). Their work, along with that of others, found that people do not judge uncertainty according to the rules of probability and statistics. This is due, in the main, to their using heuristics that in turn lead to cognitive biases and limitations, some of which are discussed next.

Cognitive Suboptimality

A common way to demonstrate cognitive suboptimality in the laboratory has been to define a judgment or choice task, determine the optimal response, usually by comparing it with one obtained by applying a Bayesian model (Slovic & Lichtenstein, 1971; see also Edwards, 1968; Phillips & Edwards, 1966), and to observe the extent to which actual behavior deviates from the optimal response. Much of this research focuses on the limitations and errors of probabilistic thinking and is pessimistic about cognition. It also contains explanations for, and possible solutions to, these limitations. I illustrate these issues by arbitrarily discussing four biasing or limiting phenomena (out of a possible dozen or more). More complete reviews of these can be found in the so-called judgment and decision literature (e.g., Arkes & Hammond, 1986; R. H. Ashton et al., 1989; Einhorn & Hogarth, 1981; Hogarth, 1987; Kahneman, Slovic, & Tversky, 1982; D. N. Kleinmuntz, 1987; Pitz & Sachs, 1984; Slovic, Fischhoff, & Lichtenstein, 1977; Wright, 1985). The four phenomena are *illusory correlation, overconfidence, relevance of experience,* and *cognitive overload.*

The first of these, uncovered by Chapman and Chapman (1969), demonstrated how people's prior expectations of perceived relations bias inferences. Chapman and Chapman taught naive subjects to associate personality characteristics with human figure drawing cues. Most subjects indeed learned to see what they expected to see. The subjects also overestimated the frequency of the learned cooccurrences. One example is the association of large eyes with suspiciousness, an illusory correlation that has also become popular among experienced clinical psychologists (Chapman & Chapman, 1971; for a critique of these studies, see Hammond, 1986).

Another biasing phenomenon is overconfidence, which was identified by Oskamp (1962, 1965) among clinical psychologists whose confidence, but not necessarily diagnostic accuracy, increased when provided additional information about psychiatric cases. Worse yet, Holsopple and Phelan (1954) earlier reported that the most confident clinicians tend to be the least accurate (see also Arkes, 1981), a finding that has serious implications for unwary patients.

Generally, overconfidence can be assessed by keeping a box score of the frequency of predicted outcomes relative to their actual occurrence. According to Einhorn (1980b), this helps "calibrate" human judges by disclosing their forecasting accuracy. Unfortunately, very few experts bother to keep tallies of their performance, perhaps because of sloth, poor record-keeping habits, or, more likely, because they lack awareness of their past cognitive strategies. Evidence of this lack was found by Huesmann, Gruder, and Dorst (1987). They demonstrated the inability of subjects to report hypnotically induced "forbidden" search strategies in memory. Yet subjects still used this information to solve problems. Similarly, Lewicki, Hill, and Bizot (1988) showed that although unconsciously acquired knowledge can facilitate performance, it cannot always be articulated. Hence, if people will not tally or cannot explicate their forecasting strategies, they cannot reconstruct them. Nor are recall and explication necessarily accurate (see Ericsson & Simon, 1980; Nisbett & Wilson, 1977).

Typical findings are that overconfident people overestimate how much they know, even about the easiest knowledge tasks. They are reasonably well calibrated when their announced odds are low (1:1–3:1), but are less so when their publicized odds are high. Furthermore, they are well calibrated for sports and weather forecasting (for a review of the forecasting literature, see Fischhoff & MacGregor, 1982, 1986; Lichtenstein & Fischhoff, 1977; MacGregor & Slovic, 1986; Murphy & Winkler, 1984; Pitz, 1974; Ronis & Yates, 1987; Yates & Curley, 1985; see also Fischhoff, Slovic, & Lichtenstein, 1977; Lichtenstein, Fischhoff, & Phillips, 1977, 1982; and Wright & Ayton, 1986). Most findings suggest that people tend to be overconfident and thus poorly calibrated for some events on some occasions but are well calibrated under certain circumstances and for some events. Other current research and discussions aimed at sorting out these events, occasions, and conditions appear in a wide variety of domains (e.g., Bazerman, 1983; Edwards & von Winterfeldt, 1986; Saks & Kidd, 1980; Slovic, 1982; Thaler, 1986).

The problem for most people is that overconfidence, as noted by Einhorn and Hogarth (1978), leads to overweighting of the importance of occurrences that confirm their hypotheses. This results in their ignoring or not collecting information that may be unfavorable to their hypotheses. This, in turn, impedes learning from environmental feedback, with its deleterious effect on future predictions (see Einhorn, 1980a; Goldberg, 1968b; see also Einhorn & Hogarth, 1981, and Hammond et al., 1973).

Regarding the relevance of experience for predictive accuracy, research suggests that experience alone may not be important (Garb, 1989; Goldberg, 1959, 1968b, 1970; B. Kleinmuntz & Szucko, 1982, 1984, 1987; Oskamp, 1962, 1965; Szucko & Kleinmuntz, 1981; Turner, 1966; Watson, 1967). In many decision settings, inexperienced practitioners, and even naive laboratory subjects, perform as well (or as poorly) as more experienced ones (Goldberg, 1959). These results, according to Brehmer (1980), are exactly what they should be, given that experience alone often yields little feedback information from which to learn (for an extensive review of the effects of cognitive feedback on multiple measures of performance, see Balzer, Doherty, & O'Connor, 1989).

Finally, research that has its roots in the information-processing psychology of Newell and Simon (1972) provides additional evidence suggesting that cognition is bounded. For example, studies by Kotovsky, Hayes, and Simon (1985) found that memory capacity and cognitive processing capability in solving toy problems (e.g., Tower of Hanoi) are easily overloaded. Correct problem solution depends on learning to use appropriate decision rules for the problem at hand, which, in turn, calls for careful study of the features of complex problems *and* the capacities of the information processor.

Cognitive Correctives

The emerging judgment and decision literature is attending increasingly to *debiasing,* which is aimed at identifying variables that contribute to poor judgment. By so doing, the hope is to control and eliminate systematic bias. Fischhoff (1980, 1982a, 1982b, 1987), for example, divided the putative biasing culprits as follows: those due to faulty tasks, those due to faulty judges, or mismatches between judges and tasks.

Regarding biases due to faulty tasks, Fischhoff (1980, 1982a) noted that experimenters possibly present subjects with unfair tasks (e.g., subjects did not care about, were confused by, became suspicious of, were unable to express what they knew about, or were given too many tasks). They also present confusing or carelessly designed tasks that overlook what subjects can or cannot do. Fischhoff's solutions are to clarify task instructions, use better response modes, and ask fewer questions.

Biases that arise because of faulty judges are traceable to the selection of subjects who are incorrigibly untrainable. This can be corrected in part by extended training programs with feedback. It would be best, however, not to select such subjects. When neither the task nor the judge is apparently at fault, Fischhoff (1980, 1982a) called for an examination of the person–task situation. He suggested selecting subjects with domain specific expertise or restructuring tasks that permit the best use of existing cognitive skills.

In this context, Nisbett, Krantz, Jepson, and Kunda (1983) recommended the use of formal training in statistics as a corrective. Their recommendation originates in research showing that everyday inductive reasoning is roughly equivalent to using formal statistical principles. Similar advice is offered by others interested in having experts avoid common judgmental biases because once a bias is in place, its influence is difficult to control (e.g., see also Arkes, 1981; Chapman & Chapman, 1969; Christensen & Elstein, in press; Fischhoff, 1979; Politser, 1987; Wood, 1978).

From among these debiasing solutions, the easiest to implement is that of clarifying instructions to subjects. It may also be the most useful. For example, Svenson (1985) found that, among undergraduates challenged by complex laboratory judgment tasks designed to elicit their estimates of probable death risks of persons depicted in eight hypothetical cases, confusion over task requirements caused risk overestimation. Svenson also reported that systematic risk overestimation occurred when they failed to incorporate fully the relevant instructions. On the other hand, people who understood the task attained proper approximations of risk, a finding also reported by Dod (1988) in a study of physicians' risk preferences.

Hogarth (1981) has taken another approach in this regard. His contention is that behavioral decision research needs to focus on continuous prediction occurring in dynamic and complex task environments (see also Hogarth & Makridakis, 1981; Neisser, 1976). By adopting such a framework, laboratory researchers would more closely approximate real-world decision making. Using a simulated continuous and dynamic laboratory task environment, D. N. Kleinmuntz (1985, 1987) and D. N. Kleinmuntz and Thomas (1987) did indeed demonstrate that the use of this framework can lead to new insights about decision making.

Using a different research paradigm, Hammond, Hamm, Grassia, and Pearson (1987) proposed that there is a time for pure intuition and a time for quasi-rational and analytical reasoning. When to use which depends largely on a problem's task characteristics. Accordingly, Hammond et al. devised a cognitive continuum, ranging from intuition at one end to analysis at the other. They also conceptualized a corresponding range of task conditions. At the intuitive pole, the tasks require rapid, unconscious data processing that combines available information by simple averaging. It has low reliability but is moderately accurate. Analysis, the other end of the continuum, is relatively more slow, conscious, and deliberate, but its reliability and accuracy are higher. It entails aggregating information by using organizing principles that are more complicated than averaging. Hammond et al. (1987) do not claim that analytical reasoning is without error. It can produce extreme error. Nevertheless, the importance of their schema, which they tested empirically among a group of highway engineers, is that it permits detailed analyses of how error arises. They did so by facilitating comparisons of intuitive, quasi-rational, and analytical cognition under several task conditions and by closely adjusting the correspondence between the type of task presented (intuition inducing vs. analysis inducing) and the cognitive activity selected (intuition vs. analysis).

To summarize this review so far, I have argued that people are indeed not as good as they think they are at using their heads, but that they can be debiased in a variety of ways, can be formally trained to minimize error, and can be guided to make better decisions. Moreover, by modifying experimental environments so as to resemble real-world complex tasks, people's reasoning can be improved. Finally, by means of detailed analyses of when to use intuition versus more analytical thinking, people can reduce judgmental error.

Having summed up the discussion in this way, it is also noteworthy that there are some who argue for the limited generalizability of laboratory research to real-world decisions (e.g., Christensen-Szalanski & Beach, 1984; Christensen-Szalanski & Bushyhead, 1981; Ebbesen & Konecni, 1975, 1980; Fischhoff, 1987; Funder, 1987). Others, however, have shown that people can be adaptive even in laboratory settings (e.g., Klayman, 1984, 1988; Klayman & Ha, 1987; Paquette & Kida, 1988; Payne, Bettman, & Johnson, 1988; Reder, 1987; see also Payne, 1982). Still others have argued that although people can be stupid in experimental rooms, they function quite adequately in a cognitively complex world (e.g., Toda, 1962; see also Toda, 1980). The question now, given that rational reasoning is possible with proper training and under some conditions, is whether decision aids can improve thinking.

Aided Judgment

So far I have noted that judgment can be outperformed by simple linear composites of predictors and how decisions can be modeled by a variety of regression approaches. This section focuses on three other types of formula. These are, in turn, the Bayesian, signal detection, and computer approaches.

Bayesian View

An alternative to the algebraic and additive correlational schemes already encountered is the Bayesian paradigm. It has its recent roots in Savage's (1954, 1972) work on statistical decision theory. Savage was a pioneer in promoting the idea of formalizing subjective probability by combining data and beliefs about the data (see also Edwards, 1954, 1961, 1962, 1971; Edwards, Lindman, & Savage, 1963; Meehl & Rosen, 1955; Slovic & Lichtenstein, 1971). The Bayesian approach can help optimize predictions under uncertainty. It does so by offering a normative model of how people *should* think if they are to think optimally.

A well-known early example of its use in clinical psychology grew out of Rosen's (1954) psychometric efforts and difficulties at predicting suicide, an infrequent occurrence even among the psychiatrically hospitalized. Meehl and Rosen (1955), in a subsequent analysis, demonstrated that in predicting such rare events it is helpful to apply Bayes's rule. Their analysis showed that by incorporating appropriate base rates (i.e., prior probabilities), the Bayesian formula improves on unaided intuition. Moreover, Meehl and Rosen argued that a psychometric device, to be efficient, must outperform predictions based only on prior probability data—a seemingly obvious point, but one that is counterintuitive to practicing clinicians (see also Rorer & Dawes, 1982, on bootstrapping psychometric base rates and Grove, 1985, on why this procedure is not cost efficient in bootstrapping diagnoses).

Decision Analysis

A formal technique that incorporates Bayes's theorem, decision analysis, is a more recent and elaborate decision support procedure. It adds two essential components to conventional Bayesian thinking (e.g., Edwards, 1971, 1977; Edwards et al., 1963; Hogarth, 1987, especially pp. 177–203; Howard, 1966; Keeney, 1982; Keeney & Raiffa, 1976; Raiffa, 1968). Stated here as questions, these are as follows: (a) What are the consequences of alternative actions? and (b) what are the uncertainties in the environment relevant to the actions and their consequences?

Decision analysis as a decision aid has been applied in a variety of nonclinical (e.g., Arkes & Hammond, 1986, pp. 4–7; Bell, Keeney, & Raiffa, 1977; Gardiner & Edwards, 1975; Kaufman & Thomas, 1977; Keeney, 1982; von Winterfeldt & Edwards, 1986) and clinical disciplines (e.g., Beck, 1986; Beck & Pauker, 1983; Pauker & Kassirer, 1987; Sisson, Schoomaker, & Ross, 1976; for critical and technical reviews, see Hershey & Baron, 1987; Hogarth, 1987, pp. 177–184; Politser, 1981, 1984; Politser & Fineberg, 1987). Decision analysis is useful because it decomposes complex problems, thus simplifying them. It often depicts graphically, in the form of decision trees, the courses of action open to the decision maker, the probabilities associated with their outcomes, and their corresponding consequences. These components are aggregated multiplicatively. Assuming the technique's correct application, decision analysis assists in making better decisions.

A critical assumption in applying this technique, of course, is that experts' preferences, beliefs, and likelihood functions are elicited accurately. Or, as Pitz (1974) reminded readers, "in any decision analysis, the [subjective] evaluation of uncertainty at each stage of the decision is critical to the final solution" (p. 41). The procedure is admittedly not free of subjectivity and, hence, of error.

According to D. N. Kleinmuntz (1990), it is important to identify and control such error potential (e.g., Fischhoff, 1980, 1982a; Hogarth, 1975; Lichtenstein et al., 1982; Wallsten & Budescu, 1983). It is also important to assess the accuracy of preference elicitation (Farquhar, 1984; Fischhoff, Slovic, & Lichtenstein, 1980; Hershey, Kunreuther, & Schoemaker, 1982). D. N. Kleinmuntz (1990) listed numerous corrective procedures that can then be applied to reduce the effects of error in decision-analytic models. The correctives include (a) using multiple assessments to check for consistency, (b) performing sensitivity analyses to modify probabilities and preferences for specific decisions, and (c) building error theories designed to predict, explain, and control the cumulative impact of error on inference and judgment.

Signal Detectability

Signal detection research, also an outgrowth of statistical decision theory, was originally developed to help detect radar signals in air traffic control systems (e.g., Tanner & Swets, 1954). The idea was to evaluate observers' ability to detect simple sensory stimuli embedded in noise. A practical example would be a situation in which airplane "blips" must be identified on a radar screen. The signals, in this case, the blips, are observed against a background of noisy or extraneous echoes. The observation task is to discriminate between the two classes of events, signals and noise, a seemingly trivial task.

Decision accuracy, however, depends on the *decision criterion* used by observers. The criterion is influenced by cognitive threshold limitations and by various random and systematic biases that, in turn, cause deviations from discriminative optimality. Signal detection theory provides a normative standard with which to compare the precision of observers with their empirical performance. The technique can thus be used to improve detection performance because the comparison yields values that disclose differences between the inherent detectability of noisy signals and the ability to detect them. It has been used successfully in a wide variety of settings where an individual's or a diagnostic system's predictive accuracy must be evaluated (e.g., Lopes, 1982; Mowen & Linder, 1979; Swets, 1964, 1986, 1988; Szucko & Kleinmuntz, 1981, 1985).

Information-Processing View

Unlike much of the judgment and decision research discussed so far, which has its roots in correlational or statistical decision theory, information-processing psychology views computer program statements as formalisms to represent intelligent problem solving. Its goal is to construct by modeling cognition what McCorduck (1979) called "machines who think." This view broadens the definition of formula to include more than mathematical or statistical approaches as decision support systems. It probably comes closest to Meehl's (1954, p. 38) idea that computers may someday replace thinking.

The computer's potential as a surrogate intelligent system, according to this view, is that its software statements can be used as elements of psychological theories. The idea of computer thinking received an important impetus from the work of Newell and Simon (1961, 1972). Over the years, they have argued and demonstrated (with several generations of information-processing languages) that people can articulate their reasoning by producing *thinking-aloud* protocols while solving problems. Computer thinking, however, has not yet realized its potential (see Reynolds, 1987, pp. 12–13, for a brief description of computer uses in psychology and computer science).

Perhaps the most important outgrowth of this view to date has been to cast the computer into a new and important role. For example, it has led to artificial intelligence, or AI, and expert systems research, which offers the possibility of the computer as a decision support. This expert systems use is described further in the next section on aided and unaided intuition. It will suffice here to indicate that it shares with other formulas the possibility of becoming a powerful tool for aiding thinking, and to note that expert systems are product-directed computer programs, whereas research in AI, generally, is more theory-directed (see Schank, 1984, pp. 32–38, for a clarification of this distinction).

To summarize this section on aided judgment, it is apparent that Bayesian and decision-analysis approaches can offer valuable decision supports if they are properly applied. The errors that arise are often identifiable and therefore controllable. Another decision aid, signal detectability, can also help evaluate and augment judgmental accuracy. Information-processing psychology, which proposes a descriptive modeling approach to thinking, can do likewise. It is also important to note that in applying these formulas, intuitive inputs and monitoring are essential. Therefore, it seems wise to consider the possibility of using the best of both approaches: decision support and judgment, as indicated in the following section.

Statistical *and* Intuitive Judgment

Some years ago, Edwards (1962) and then Sawyer (1966) proposed that experts can contribute to predictive inference by providing judgments that could be aggregated mechanically. Following up on this idea, Einhorn (1972) demonstrated that when expert measurement and the formula are used together, the combination outperforms either method used alone. For example, he studied expert pathologists who predicted cancer survival. He found that their predictive accuracy was improved by using their heads as measuring devices and formulas as rules to combine the measurements.

Similarly, Blattberg and Hoch (in press) have shown this in a managerial context. In five different business forecasting situations, a 50% model plus a 50% manager solution outperformed either of these decision modes in isolation. Evidently, the improvement over unaided judgment was due to the formula's capitalizing on both the intuiter's "special case" insights (i.e., as in Meehl's, 1954, broken leg example) and the model's reliable combination of this information. A similar result was reported by Showers and Chakrin (1981) in revenue collection. They used the formula as a customer credit screen and the head to provide inputs to their credit evaluation procedures.

Yet another aggregating tack has been reported by Hogarth (1978). He proposed that the validity of expert judgment is enhanced by forming *staticized* groups (i.e., aggregating the opinions of two or more experts). Thus, he developed an analytical model that, given certain conditions, yields group validity data that suggest how many experts should be included in a staticized group. It can also help decide which expert(s) may be added or deleted in order to attain optimality. This and similar group models of judgment and problem solving have been tested in a variety of laboratory and real-world settings and have been found to be quite efficient at improving decision making (e.g., A. H. Ashton & Ashton, 1985; R. H. Ashton, 1986; Clemen, 1986; Clemen & Winkler, 1985, 1987; Davis, 1969; Hill, 1982; Libby & Blashfield, 1978; Makridakis & Winkler, 1983; Morris, 1983, 1986; Steiner, 1972; Winkler, 1986; Winkler & Makridakis, 1983).

Bootstrapping (see Dawes, 1971; Dawes & Corrigan, 1974) also provides an illustration of the combined use of heads and formulas. It is the phenomenon whereby a model of the person or persons outperforms the unaided intuition of the modeled person or persons. Essentially, its rationale is quite simple. Humans provide predictor inputs, assigning them their putative weights and monitoring the directions of the resulting predictions. The formula's contribution is its consistent decision-rule application and integration. This head–formula combination can work well (e.g., Bowman, 1963; Camerer, 1981; Dawes, 1971, 1988; Dawes & Corrigan, 1974; Hammond, 1955; Hoffman, 1960, 1968; Hogarth, 1978; B. Kleinmuntz, 1963). But there are problems.

One problem, as Slovic (1972) noted, is that modeling intuition can preserve and reinforce, and perhaps even magnify, existing cognitive biases. The assumption of bootstrapping, however, as well as that of most mechanical processing techniques, is that despite its inclusion of cognitive biases, the prediction formula invariably outperforms unaided intuition because the increased reliability attained by the formula outweighs any effects of bias and intuition (Robin Hogarth, personal communication, February 19, 1989).

Another problem with bootstrapping is that the judges being modeled may not be cognitively competent. B. Kleinmuntz and

Szucko (1987), for example, found this to be the case among polygraphers thinking aloud while analyzing lie detection protocols. The difficulty was that the polygraphers were highly fallible. Their predictions were the equivalent of a crapshoot. Even so, as Arkes (personal communication, October 6, 1988) has indicated, the modeling of their reasoning should surpass their performance so long as they provide judgments with even an iota of validity. So it did in several earlier studies where it was shown with lens modeling and signal detection theory how and why polygraphers do not optimize their predictions with the information provided them (B. Kleinmuntz & Szucko, 1982, 1984; Szucko & Kleinmuntz, 1981, 1985).

Cognitive inputs are also important when using Bayes's formula. This is best articulated by such Bayesians as Savage (1972) and Edwards (1972). For example, Savage advocated the use of formal inference for medical diagnosis in combination with the human's "wonderful abilities to make such informal diagnoses, *for which there is sometimes no formal substitute yet available*—as when we recognize an odor or a face" (p. 134; emphasis added). In a similar vein, Edwards (1972) stated, "there are actually two intellectual steps in diagnosis after data collection is complete. One is the *judgment of the meaning* of each individual symptom; the other is the aggregation of the symptoms to reach a diagnosis" (p. 140–141; see also Berger & Berry, 1988, and Dawes, 1988, on the importance of human judges). The italicized statements are intended to emphasize the importance of subjectivity even in formal procedures.

The entry of computers as decision supports is a relatively recent phenomenon. This expert systems use of computers is designed to perform highly specialized knowledge tasks. Toward this end, so-called knowledge engineers provide strategies and information to computers, by eliciting it either from experts or from textbooks, or both (e.g., Barr & Feigenbaum, 1981; Newell & Simon, 1972; Schank, 1984; Simon, 1979; Waterman, 1986).

Expert systems have been found to be especially useful in fields with shortages of qualified specialists. Their outperformance of humans is due to their ability to accumulate, organize, and codify large quantities of knowledge. Expert systems also decompose, formulate, and view new problems so that they are easy to solve. They do so by searching through a set of possible solutions, finding an efficient or acceptable one, and then modifying and permanently storing the engineered expertise. Once the knowledge and strategies have been acquired and stored, the computer becomes a reliable and swift aggregating tool.

Recent examples of specialties in which expert systems have been constructed as decision supports include analogical problem solving (Eliot, 1986), outer-space-station operation (Leinweber, 1987), legal reasoning (Wiehl, 1989), oncology protocol management (Shortliffe, 1986, 1987), general medical diagnosis (Barnett, Cimino, Hupp, & Hoffer, 1987), and emergency room prediction of myocardial infarction among chest pain patients (Goldman et al., 1988). Most of these systems use a combination of the expert's knowledge and other available information about how best to solve the problem at hand. Current work in the area focuses on doing a task analysis, in combination with the expert, and then designing a system that performs such tasks well. Edwards (personal communication, March 9, 1989) called these *competent systems* to suggest that such task-analysis-based systems can outperform experts, not just simulate them.

Benefits and Costs of Combining Heads With Formulas

Given the apparent success of diverse formulas, when used alone or together with intuition, and considering the many pitfalls of unaided intuition, one may well ask, "Why are we still using our heads instead of formulas?" The answers to this query are many, depending on whom one asks.

Deluded Self-Confidence

A partial answer can be found in the example of de Dombal's research in internal medicine. Over a period of more than a dozen years, he and his coworkers (de Dombal, 1984b; de Dombal, Horrocks, & Walensley, 1975; and de Dombal, Leaper, Stanilaud, McCann, & Horrocks, 1972) developed computer-based Bayesian diagnostic systems of acute abdominal pain. These were quite successful in that by 1975, one of the system's early versions reached 91% accuracy, outperformed senior clinicians (de Dombal et al., 1975), and proved to be largely generalizable, especially when the prior probabilities were properly adjusted to match the local population. Its advantages over unaided judgment were clear.

Despite these virtues, however, de Dombal (1984a) did not recommend the system's routine use. He expressed reservations about its first-rate performance, particularly because human well-being was at risk. In such high-risk situations, he felt, decision makers should rely on their own, not a computer's, expertise. This self-confidence in human expertise has been confirmed in a laboratory study. Arkes, Dawes, and Christensen (1986), for example, have shown that the acceptability to users of a decision aid does not rest on whether it substantially outperforms unaided judgment. Rather, it depends on their *belief* that they have *real expertise* in a domain, thus inspiring confidence in the possibility of beating the odds. That such is the case has subsequently been demonstrated by de Dombal, who has written that he and his group "showed that throughout the UK doctors' performance levels were poor and could themselves improve with the aid of a computer-based decision support [and] a queue formed" (de Dombal, personal communication, March 21, 1989; see also Adams et al., 1986).

Configural Complexities

Another argument often heard in favor of heads over formulas, particularly in clinical settings, is that clinical decisions entail integrating complex *patterns* of symptoms and signs. Presumably, these are due to their task environments being more ill-structured than those of game playing (e.g., see Abelson, 1985; Wilkinson, Gimbel, & Koepke, 1982). Meehl (1967) called these patterns "configurated functions" that are "visual gestalten [that] can be perceived without the percipient's knowing the underlying formula" (p. 597).

Complexity, however, also characterizes the problems encountered in game playing. Yet game-playing computer programs have been somewhat more successful and have met with greater acceptance than clinicians' computer programs. For example, Simon (1979) estimated that a good chess player needs to know some 1,300 chess piece and position patterns in order to play well; masters and grandmasters, some 50,000 or more such configurations. These estimates of the number of patterns and their complexity *seem* on a par with those existing in many cognitive tasks confronting clinicians. Why, then, have computers not been equally successful among clinical specialists?

Probably because the analogy between clinical information processing and game playing can be pushed too far. There are some similarities, but there are many more differences. Compared with the problems encountered in the clinical sciences, games are well structured and the moves and rules are clearly defined. Moreover, less is at risk if one loses. Clinical problems, *per contra,* are ill-structured. The "opponent" is nature, which carries with it more uncertainty and ambiguity than chess and other games. Also, as noted in the de Dombal example given earlier, the stakes are high if one loses, thus rendering the comparison with chess tenuous. What is needed in clinical settings, however, to carry the analogy further, is an international grandmaster program, one that beats the odds most of the time. Such convincing evidence may persuade even the most self-confident clinicians to abandon the use of unaided intuition.

Costs and Availability of Decision Supports

Yet another reason for not using formulas with or without intuition is the costs. These can be computed, but not easily. One way is to compute the added error costs incurred in any of the three modes (i.e., formula vs. intuition vs. both used together). The time and effort invested in doing so can be considerable. Few people possess the financial resources or level of technical and experimental sophistication needed to test the quality of decision making.

Regarding the error possibilities themselves in using formulas, Fischhoff and Beyth-Marom (1983) noted that even Bayesian inference can be error-prone in several ways. First, judgmental and other cognitive biases and miscalculations can disrupt its proper application. Second, one can formulate the wrong hypotheses for particular predictions or actions to be taken. Third, as already noted earlier, one may err in eliciting beliefs and values before incorporating these into a decision analysis. Fourth, prior probabilities or likelihood functions can be estimated or observed incorrectly, or may be ignored altogether. Fifth, even if all the foregoing procedures are correct, one can use the wrong aggregation rule (i.e., averaging instead of multiplying) or apply the right one incorrectly. An evaluation of any of these possibilities can be time and labor intensive.

The other costs of using decision aids can occur in deciding which to use and when and how to use them, assuming one is available. Even here, people need to exercise good judgment; or, to phrase this in decision-analysis terms, the decision to use an aid is a large and difficult choice problem that must itself be decomposed. Thus, the use of an aid, including the decision to use one, requires a high degree of technical wherewithal about the assumptions underlying their proper application. This, too, can be costly, in terms of both time and money (e.g., see Fischhoff, 1980; Hogarth, 1987, p. 197; D. N. Kleinmuntz, 1987).

All of these cost considerations assume that decision support systems are readily available and appealing to prospective users, although most decision makers are unaware of their availability. When made aware of them, they may not use, or may even oppose, their implementation. Clinicians, for instance, or engineers, managers, and others considering the use of decision supports, are unfamiliar with many of the techniques discussed in this article. They need to have had contact with the literature on linear models, decision analysis, signal detection, or expert systems. Only a very few have. Even if they have read about these esoteric aids, according to Hogarth (1987, p. 199) "there are a number of resistances to such quantification." These resistances may be quite irrational and based on egocentric and emotional grounds. Moreover, only a few of the knowledgable users have the time or ability to design experiments to evaluate the decision support system's efficacy. This is, again, a difficult and costly undertaking. The dilemma leaves people no choice but to use their heads in deciding whether to use any system at all. In these instances, they usually end up using their heads instead of formulas.

Thus, in sum, one can see that despite the possible advantages of using decision supports, their implementation is often difficult. The difficulty is sometimes monetary, temporal, emotional, or technical. More often it is in the form of the unavailability of the aids and the means for their evaluation. Outside consultation, if affordable, may be necessary; but that, too, is often unavailable.

Summary and Concluding Comments

The answer as to why people still use their heads, flawed as they may be, instead of formulas, is that for many decisions, inferences, choices, and problems there are as yet no available formulas. When formulas are available, their evaluation is not feasible, when used either alone or in combination with intuition. Coming full circle to the recognition that people may have to use their heads *instead* of, or *together* with, formulas while awaiting new decision support developments, I offer the following guidelines. All of these emerge from the review, but were not previously explicated as such. Meanwhile, the reader may take comfort in Payne et al.'s (1988; Payne, Bettman, & Johnson, 1990) findings, which show that although people's decisions are sometimes suboptimal, they can adapt in directions representing optimal efficiency–accuracy trade-offs:

1. People could delineate the types of decisions that do not easily lend themselves to intuition. For example, it is counterproductive to compute and assign optimal weights to cues. Nor should one attempt to apply decision rules in one's head. In these instances, it is advisable to use a calculator or an appropriate aggregating formula. Recall here Meehl's (1986) advice "not to eyeball the heap of purchases" (p. 372) at the supermarket checkout counter; just add it up.

2. Likewise, one could identify the types of decision problems that are not readily formalized. Meehl (1967), for example, noted that these include predictions that are open ended (i.e., the content of the criterion is created rather than prespecified) and that deal with special cases. Other types include inferences that require as yet unarticulated decision rules or those

that necessitate the intuitive development of a theory of the phenomena under observation. There are also judgments that need to be made or solved quickly if they are to be practical. Add to these the use of the head when available formulas are as yet unevaluated or unvalidated and when dealing with special cases as suggested by Meehl's (1954) broken leg example.

3. Using Guidelines 1 and 2, one could develop meta-rules that stipulate when to use the formulas, heads, or a mixture of both. This calls for familiarization with most available decision aids, as well as with situations for their use. An example of a meta-rule might be to use Blattberg and Hoch's (in press) equal weighting combination of 50% model and 50% expert. Further research may show other solutions to be more appropriate for specific classes of problems. Meanwhile, the 50:50 solution is appealing, first because it is simple. Second, it overcomes some of the resistance to decision supports because it provides experts with the opportunity to continue to participate in decisions. Finally, it has the demonstrated advantage of being invariably more accurate than using either formulas or heads alone.

4. It may be helpful, as well, to differentiate between types of decisions that involve backward rather than forward reasoning. Most clinical inference, for example, is characterized by backward reasoning in that diagnosticians often attempt to link observed effects to prior causes. Compared with this form of post hoc explanation, statistical prediction entails forward reasoning because it is concerned with forecasting future outcomes, given observed information. Whereas the former provides decision makers with many degrees of freedom, statistical reasoning soon confronts them with discrepancies between predicted and actual outcomes. Furthermore, clinical and statistical approaches rest on different assumptions about random error. The clinical side considers error a nuisance variable. The statistical approach, *per contra,* accepts error as inevitable, and in so doing makes less error in prediction in the long run (Einhorn, 1986; see also Einhorn, 1988; Einhorn & Hogarth, 1982, 1986).

5. Regardless of whether backward or forward inferences are made formally, intuitively, or both, one should record their accuracy during the course of a day, week, month, and so on. This entails careful documentation of the heuristics or formulas used for decisions—not an easy assignment. Then what is one to do? Garb's (1989) meta-rule is that when such box-score tallying is not feasible, or when unbiased feedback is unavailable, experts should use available decision aids rather than intuition.

6. Extreme confidence in one's predictive accuracy, particularly without the benefit of outcome feedback, should be a red flag suggesting that, in all likelihood, the predictions are flawed. A post hoc analysis of these predictions is advisable and may indeed identify where one has erred.

7. Considering that people tend to seek out confirming data once they have formulated an idea or hypothesis, researchers should follow Hogarth's (1987, p. 118) suggestion to systematically search for evidence that may disconfirm such a formulation. This is equally true for refuting a model's accuracy. It must be subjected to tests of falsifiability. Then and only then can an idea or a decision support system be refined and strengthened (G. F. Pitz, personal communication, December 5, 1988).

8. Given that intuition involves complex data processing, one should recognize that there are four ways to stray (e.g., see Hogarth, 1987, pp. 4–7): (a) selective perception, perhaps due to anticipatory biases; (b) imperfect information processing, possibly resulting from the same biases; (c) inaccurate calculations due to cognitive limitations; and (d) incorrect reconstructions of events because of biases, faulty memory, or both. An awareness of these possibilities, and the will to ferret them out, plus the determination to correct them, may lead to better decisions.

9. Because most probabilistic estimates have been shown to be systematically biased, constructing a rough scale of predictability may be helpful for a new class of events. Kahneman and Tversky (1979) have suggested that one could then check the new predictions and their outcomes against records of past ones for similar events.

10. Because formal decision aids are also error prone, one could test their before-and-after efficacy. This can be done by careful prior planning, empirical studies, requesting outside evaluative consultation, or a combination of all of these.

11. Wherever feasible, if a formula or aggregating rule is available, its cost efficiency should be tested with and without its use. If its application to simulated (i.e., Monte Carlo) or real-world data shows only slight improvement (or none), it should be discarded.

12. If the formula is an expert system, research might focus on a design that permits it to recognize its inability to solve certain decision problems. For, as Newell (cited in Wertheimer, 1985) observed, expert systems are "shallow; they don't know what they know and why they know it" (p. 29). Moreover, future research might do well also to teach expert systems to communicate their inabilities to users and, perhaps, even suggest alternative approaches to the problem at hand.

The guidelines are practical, but they hardly explain why only a select subset of clinicians and, more generally, experts are outstanding intuiters. Why is it, for instance, that some clinicians earn formidable reputations for their expertise, while others with equal training and experience do not? How come Kasparov, who retained his international grandmaster chess championship, will in all likelihood, rely on his wits the next time around? He will probably do so opposite Karpov, not a computer. The task now, clearly, is to plan research that aids lesser mortals to become first-rate experts.

Researchers seem to be on the way to reaching that goal. One measure of this is the increased interest in studying expert–novice differences (e.g., Charness, 1981; Chi, Feltovitch, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980; Murakami, 1990; Shanteau, 1988). Another is the acceptance of decision support systems, which seems to be on the rise. One dramatic instance of the latter can be found in recent newspaper reports of Carnegie Mellon's Hitech and Deep Thought chess programs. Each uses heuristics and strategies that borrow liberally from its creator (Hans Berliner and Feng-Hsiung Hsu, respectively) and can now outwit them as well as some chess grandmasters. This is an interesting bootstrapping *tour de force.* It is also a fine example of how decision aids can work together with specialists who supply their inputs and monitor their outputs. Perhaps in the near future, tournament chess players, as well as other specialists, will show up for work accompanied by their decision support assistants. The likelihood of this possibility will increase as a function of more media exposure to such decision aids. Then prospective users and researchers as well as the public at large will be inured to their appearance in hitherto unusual places. In the interim, however, people will probably continue to rely more on their heads than on formulas.

References

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97,* 129–133.

Adams, I. D., Chan, M., Clifford, P. C., Cooke, W. M., Dallos, V., de Dombal, F. T., Edwards, M. H., Hancock, D. M., Hewlett, D. J., McIntyre, N., Sommerville, P. G., Spiegelhalter, J. T., Wellwood, J., & Wilson, D. H. (1986). Computer aided diagnosis of acute abdominal pain: A multicentre study. *British Medical Journal, 293,* 800–804.

Alexander, L. B., & Kleinmuntz, B. (1962). Computer program for the Meehl-Dahlstrom MMPI profile rules. *Educational and Psychological Measurement, 22,* 193–199.

Anderson, J. R., & Reder, L. (1987). Effects of number of facts studied on recognition versus sensibility judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 355–367.

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49,* 323–330.

Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes, 37,* 93–110.

Arkes, H. R., & Hammond, K. R. (Eds.). (1986). *Judgment and decision making: An interdisciplinary reader.* New York: Cambridge University Press.

Ashton, A. H., & Ashton, R. H. (1985). Aggregating subjective forecasts: Some empirical results. *Management Science, 31,* 1499–1508.

Ashton, R. H. (1986). Combining the judgments of experts: How many and which ones? *Organizational Behavior and Human Decision Processes, 38,* 405–414.

Ashton, R. H., Kleinmuntz, D. N., Sullivan, J. B., & Tomassini, L. A. (1989). Audit decision making. In A. R. Abdel-khalik & I. Solomon (Eds.), *Research opportunities in auditing: The second decade* (pp. 95–132). Sarasota, FL: American Accounting Association, Auditing Section.

Balzer, W. K., Doherty, M. E., & O'Connor, R. Jr. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106,* 410–433.

Barnett, G. O., Cimino, J. J., Hupp, J. A., & Hoffer, E. P. (1987). DXplain: An evolving diagnostic decision-support system. *Journal of the American Medical Association, 258,* 67–74.

Barr, A., & Feigenbaum, E. A. (1981). *The handbook of artificial intelligence* (Vol. 2). Los Altos, CA: William Kaufmann.

Bazerman, M. (1983). Negotiator judgment. *American Behavioral Scientist, 27,* 211–218.

Beck, J. R. (1986). Laboratory decision science applied to chemometrics: Strategic testing of thyroid function. *Clinical Chemistry, 32,* 1701–1713.

Beck, J. R., & Pauker, S. G. (1983). The Markov process in medical prognosis. *Medical Decision Making, 3,* 419–458.

Bell, D. E., Keeney, R. L., & Raiffa, H. (1977). *Conflicting objectives in decisions.* New York: Wiley.

Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist, 76,* 159–165.

Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision making: Judgment versus choice. *Organizational Behavior and Human Decision Processes, 41,* 1–19.

Blattberg, R. C., & Hoch, S. J. (in press). Database models and managerial intuition: 50% model + 50% manager. *Management Science.*

Blois, M. S. (1980). Clinical judgment and computers. *New England Journal of Medicine, 303,* 192–197.

Bowman, E. H. (1963). Consistency and optimality in managerial decision making. *Management Science, 9,* 310–321.

Brady, D., & Rappoport, L. (1973). Policy-capturing in the field: The nuclear safeguards problem. *Organizational Behavior and Human Performance, 9,* 253–266.

Brehmer, B. (1972). Policy conflict as a function of policy differences and policy complexity. *Scandinavian Journal of Psychology, 13,* 208–221.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica, 45,* 223–241.

Brunswik, E. (1952). *The conceptual framework of psychology.* Chicago: University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory. *Psychological Review, 62,* 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Camerer, C. (1981). General conditions for the success of bootstrapping models. *Organizational Behavior and Human Performance, 27,* 411–422.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–280.

Chapman, L. J., & Chapman, J. P. (1971, November). Test results are what you think they are. *Psychology Today,* pp. 18–22.

Charness, N. (1981). Search in chess: Age and skill difference. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 467–476.

Chi, M. I. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Christensen, C., & Elstein, A. E. (in press). Informal reasoning in the medical profession. In J. F. Voss, D. N. Perkins, & J. Segal (Eds.), *Informal reasoning and education.* Hillsdale, NJ: Erlbaum.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist, 35,* 75–78.

Christensen-Szalanski, J. J. J., & Bushyhead, J. B. (1981). Physician's use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 928–935.

Clemen, R. T. (1986). Calibration and the aggregation of probabilities. *Management Science, 32,* 312–314.

Clemen, R. T., & Winkler, R. L. (1985). Limits for the precision and value of information from dependent sources. *Operations Research, 33,* 427–441.

Clemen, R. T., & Winkler, R. L. (1987). Calibrating and combining precipitation probability forecasts. In R. Vierte (Ed.), *Probability and Bayesian statistics* (pp. 97–110). New York: Plenum Press.

Davis, J. H. (1969). *Group performance.* Reading, MA: Addison-Wesley.

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist, 26,* 180–188.

Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 3–11). Hillsdale, NJ: Erlbaum.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582.

Dawes, R. M. (1988). You can't systematize human judgment: Dyslexia. In J. Dowie & A. S. Elstein (Eds.), *Professional judgment: A reader in clinical decision making* (pp. 150–162). New York: Cambridge University Press.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95–106.

de Dombal, F. T. (1984a). Clinical decision making and the computer: Consultant, expert, or just another test? *British Journal of Health Care Computing, 1*, 7–12.

de Dombal, F. T. (1984b). Future progress for computer aids as gastroenterologists. *Frontiers of Gastrointestinal Research, 7*, 186–198.

de Dombal, F. T., Horrocks, J. C., & Walensley, A. (1975). Computer-aided diagnosis and decision-making in the acute abdomen. *Journal of the College of Physicians of London, 9*, 211–223.

de Dombal, F. T., Leaper, D. J., Stanilaud, J. R., McCann, A. P., & Horrocks, J. C. (1972). Computer aided diagnosis of acute abdominal pain. *British Medical Journal, 2*, 9–18.

Dod, J. M. (1988). *Risk preferences among experienced and inexperienced physicians*. Unpublished doctoral dissertation, University of Illinois, Chicago, Illinois.

Dudycha, L. W., & Naylor, J. C. (1966). Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance, 1*, 110–128.

Ebbesen, E. B., & Konecni, V. J. (1975). Decision making and information in the courts: The setting of bail. *Journal of Personality and Social Psychology, 32*, 805–821.

Ebbesen, E. B., & Konecni, V. J. (1980). On the external validity of decision-making research: What do we know about decisions in the real world? In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 21–45). Hillsdale, NJ: Erlbaum.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380–417.

Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology, 12*, 473–498.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors, 4*, 59–73.

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

Edwards, W. (1971). Bayesian and regression models of human information processing—A myopic perspective. *Organizational Behavior and Human Performance, 6*, 639–648.

Edwards, W. (1972). N = 1. In J. A. Jacquez (Ed.), *Computer diagnosis and diagnostic methods* (pp. 139–151). Springfield, IL: Charles C Thomas.

Edwards, W. (1977). How to use multiattribute utility measurement for social decision making. *IEEE Transactions On Systems, Man and Cybernetics, SMC-7*, 326–340.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*, 193–242.

Edwards, W., & von Winterfeldt, D. (1986). On cognitive illusions and their implications. In H. Arkes & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 642–679). Cambridge, England: Cambridge University Press.

Einhorn, H. J. (1972). Expert measurement and mechanical combination. *Organizational Behavior and Human Performance, 7*, 86–106.

Einhorn, H. J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology, 57*, 562–571.

Einhorn, H. J. (1980a). Learning from experience and suboptimal rules in decision making. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 1–20). Hillsdale, NJ: Erlbaum.

Einhorn, H. J. (1980b). Overconfidence in judgment. In R. A. Shweder & D. W. Fiske (Eds.), *New directions for methodology of social and behavioral science: Fallible judgment in behavioral research* (pp. 1–16). San Francisco: Jossey-Bass.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50*, 387–395.

Einhorn, H. J. (1988). Diagnosis and causality in clinical and statistical prediction. In D. C. Turk & P. Salovey (Eds.), *Reasoning, inference and judgment in clinical psychology* (pp. 51–70). New York: Free Press.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85*, 395–406.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32*, 53–88.

Einhorn, H. J., & Hogarth, R. M. (1982). Prediction, diagnosis, and causal thinking in forecasting. *Journal of Forecasting, 1*, 23–36.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 3–19.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression *and* process tracing models of judgment. *Psychological Review, 86*, 465–485.

Eliot, L. B. (1986). Analogical problem-solving and expert systems. *IEEE Expert, 1*, 17–30.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.

Farquhar, P. H. (1984). Utility assessment methods. *Management Science, 30*, 1283–1300.

Fischhoff, B. (1979). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 349–358.

Fischhoff, B. (1980). Clinical decision analysis. *Operations Research, 28*, 28–43.

Fischhoff, B. (1982a). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). New York: Cambridge University Press.

Fischhoff, B. (1982b). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 335–351). New York: Cambridge University Press.

Fischhoff, B. (1987). Judgment and decision making. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 153–187). New York: Cambridge University Press.

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

Fischhoff, B., & MacGregor, D. (1982). Subjective confidence in forecasts. *Journal of Forecasting, 1*, 155–172.

Fischhoff, B., & MacGregor, D. (1986). Calibrating data bases. *Journal of the American Society for Information Science, 37*, 222–233.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 552–564.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1980). Knowing what you want: Measuring labile values. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 117–141). Hillsdale, NJ: Erlbaum.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101*, 95–90.

Garb, H. N. (1989). Clinical judgment, clinical training and professional experience. *Psychological Bulletin, 105*, 387–396.

Gardiner, P. C., & Edwards, W. (1975). Public values: Multiattribute utility measurement for social decision making. In M. Kaplan & S.

Schwartz (Eds.), *Human judgment and decision processes* (pp. 1-37). New York: Academic Press.

Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. *Journal of Consulting Psychology, 23,* 25-33.

Goldberg, L. R. (1965). Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs, 79*(9, Whole No. 602).

Goldberg, L. R. (1968a). Seer over sign: The first "good example?" *Journal of Experimental Research in Personality, 3,* 1688-171.

Goldberg, L. R. (1968b). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23,* 483-496.

Goldberg, L. R. (1969). The search for configural relationships in personality assessment: The diagnosis of psychosis vs. neurosis from the MMPI. *Multivariate Behavioral Research, 4,* 523-536.

Goldberg, L. R. (1970). Man versus model of man: A rationale plus some evidence for a method of improving on clinical inferences. *Psychological Bulletin, 73,* 422-432.

Goldberg, L. R. (1971). Five models of clinical judgment: An empirical comparison between linear and nonlinear representations of the human inference process. *Organizational Behavior and Human Performance, 6,* 458-479.

Goldberg, L. R. (in press). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. Grove (Eds.), *Festschrift in honor of Paul E. Meehl.* New York: Cambridge University Press.

Goldman, L., et al. (1988). A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *New England Journal of Medicine, 318,* 797-803.

Grebstein, L. (1963). Relative accuracy of actuarial prediction, experienced clinicians, and graduate students in a clinical judgment task. *Journal of Consulting Psychology, 37,* 127-132.

Grove, W. (1985). Bootstrapping diagnoses using Bayes's theorem: It's not worth the trouble. *Journal of Consulting and Clinical Psychology, 53,* 261-263.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62,* 255-262.

Hammond, K. R. (1965). New directions in research in conflict resolution. *Journal of Social Issues, 21,* 44-66.

Hammond, K. R. (1978). Toward increasing competence of thought in public policy formation. In K. R. Hammond (Ed.), *Judgment and decision in public policy formation, AAAS selected symposium* (pp. 11-32). Boulder, CO: Westview Press.

Hammond, K. R. (1986). Generalization in operational contexts: What does it mean? Can it be done? *IEEE Transactions on Systems, Man and Cybernetics, SMC-16,* 428-433.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytic cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-17,* 753-770.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71,* 438-456.

Hammond, K. R., Stewart, T. R., Brehmer, R., & Steinmann, D. O. (1975). Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes* (pp. 271-312). New York: Academic Press.

Hammond, K. R., & Summers, D. A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review, 72,* 215-224.

Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome feedback on multiple-cue probability learning. *Organizational Behavior and Human Performance, 9,* 30-34.

Hastie, R., Penrod, S. D., & Pennington, N. (1983). *Inside the jury.* Cambridge, MA: Harvard University Press.

Hershey, J. C., & Baron, J. (1987). Clinical reasoning and cognitive processes. *Medical Decision Making, 7,* 203-211.

Hershey, J. C., Kunreuther, H. C., & Schoemaker, P. J. H. (1982). Sources of bias in assessment procedures for utility functions. *Management Science, 28,* 936-954.

Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 91,* 517-539.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57,* 116-131.

Hoffman, P. J. (1968). Cue-consistency and configurality in human judgment. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 53-90). New York: Wiley.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association, 70,* 271-289.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance, 21,* 40-46.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90,* 197-217.

Hogarth, R. M. (1987). *Judgement and choice: The psychology of decision* (2nd ed.). Chichester, England: Wiley.

Hogarth, R. M., & Makridakis, S. (1981). The value of decision making in a complex environment: An experimental approach. *Management Science, 27,* 92-107.

Holsopple, J. G., & Phelan, J. G. (1954). The skills of clinicians in analysis of projective tests. *Journal of Clinical Psychology, 10,* 307-320.

Holt, R. R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology, 56,* 1-12.

Holt, R. R. (1970). Yet another look at clinical and statistical prediction: Or is clinical psychology worthwhile? *American Psychologist, 25,* 337-349.

Holt, R. R. (1978). *Methods in clinical psychology: Predictions and research* (Vol. 2). New York: Plenum Press.

Holt, R. R. (1986). Clinical and statistical prediction: A retrospective and would-be integrative perspective. *Journal of Personality Assessment, 50,* 376-386.

Howard, R. A. (1966). Decision analysis: Applied decision theory. In S. B. Hertz & J. Melese (Eds.), *Proceedings of the Fourth International Conference on Operations Research* (pp. 55-71). New York: Wiley.

Huesmann, L. R., Gruder, C. L., & Dorst, G. (1987). A process model of posthypnotic amnesia. *Cognitive Psychology, 19,* 33-62.

Jungermann, H. (1986). Two camps on rationality. In H. Arkes & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 627-641). New York: Cambridge University Press.

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47,* 263-291.

Kaufman, G. M., & Thomas, H. (Eds.). (1977). *Modern decision analysis: Selected readings.* Middlesex, England: Penguin.

Keeney, R. L. (1982). Decision analysis: An overview. *Operations Research, 30,* 803-838.

Keeney, R. L., & Raiffa, H. (1976). *Decision with multiple objectives: Preferences and value tradeoffs.* New York: Wiley.

Klayman, J. (1984). Learning from feedback in probabilistic environments. *Acta Psychologica, 56,* 81-92.

Klayman, J. (1988). Cue discovery in probabilistic environments: Uncertainty and experimentation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14,* 317-330.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and hypothesis testing. *Psychological Review, 94,* 211–228.

Kleinmuntz, B. (1963). Personality test interpretation by digital computer. *Science, 139,* 416–418.

Kleinmuntz, B. (in press). Computerized clinicians: Developments and problems. In D. Cicchetti & W. Grove (Eds.), *Festschrift in honor of Paul E. Meehl.* Minneapolis: University of Minnesota Press.

Kleinmuntz, B., & Elstein, A. S. (1987). Computer modeling of clinical judgment. *Critical Reviews in Medical Informatics, 1,* 209–228.

Kleinmuntz, B., & Szucko, J. J. (1982). On the fallibility of lie detection. *Law and Society Review, 17,* 85–104.

Kleinmuntz, B., & Szucko, J. J. (1984). A field study of the fallibility of polygraphic lie detection. *Nature, 308,* 349–350.

Kleinmuntz, B., & Szucko, J. J. (1987). *Deception, lie detection, and the dynamics of legal decision making.* (Final project report submitted on Grant NES-83-19138). Washington, DC: National Science Foundation.

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science, 31,* 680–702.

Kleinmuntz, D. N. (1987). Human decision processes: Heuristics and task structure. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 123–157). North Holland: Elsevier.

Kleinmuntz, D. N. (1990). Decomposition and the control of error in decision analytic models. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 107–126). Chicago: University of Chicago Press.

Kleinmuntz, D. N., & Thomas, J. B. (1987). The value of action and inference in dynamic decision making. *Organizational Behavior and Human Decision Processes, 39,* 341–364.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17,* 248–294.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335–1342.

Leinweber, D. (1987). Expert systems in space. *IEEE Expert, 2,* 26–38.

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology, 20,* 24–37.

Libby, R., & Blashfield, R. K. (1978). Performance of a composite as a function of the number of judges. *Organizational Behavior and Human Performance, 21,* 121–129.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance, 2,* 159–183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In H. Jungermann & G. de Zeeuw (Eds.), *Decision making and change in human affairs* (pp. 275–324). Dordrecht, The Netherlands: Reidel.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (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). New York: Cambridge University Press.

Lindzey, G. (1965). Seer versus sign. *Journal of Experimental Research in Personality, 1,* 17–26.

Lopes, L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 626–636.

Lykken, D. T. (1981). *A tremor in the blood: Uses and abuses of the lie detector.* New York: McGraw-Hill.

MacGregor, D., & Slovic, P. (1986). Perceived acceptability of risk as a decision-making approach. *Risk Analysis, 6,* 245–256.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science, 29,* 987–996.

McCorduck, P. (1979). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence.* New York: W. H. Freeman.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4,* 268–273.

Meehl, P. E. (1959). A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *Journal of Counseling Psychology, 2,* 102–109.

Meehl, P. E. (1960). The cognitive activity of the clinician. *American Psychologist, 15,* 19–27.

Meehl, P. E. (1965). Seer over sign: The first good example. *Journal of Experimental Research in Personality Research, 1,* 29–32.

Meehl, P. E. (1967). What can the clinician do well? In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment* (pp. 594–599). New York: McGraw-Hill.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Meehl, P. E., & Dahlstrom, W. G. (1960). Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *Journal of Consulting Psychology, 24,* 375–387.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194–216.

Morris, P. A. (1983). An axiomatic approach to expert resolution. *Management Science, 29,* 24–32.

Morris, P. A. (1986). Observations on expert aggregation. *Management Science, 32,* 321–328.

Mowen, J. C., & Linder, D. E. (1979). Discretionary aspects of jury decision making. In L. E. Abt & I. R. Stuart (Eds.), *Social psychology and discretionary law* (pp. 219–239). New York: Van Nostrand.

Murakami, J. S. (1990). *Expert-novice differences in medical problem solving.* Unpublished master's thesis, University of Illinois at Chicago, Department of Psychology.

Murphy, A. H., & Winkler, R. L. (1984). Probability forecasting in meteorology. *Journal of the American Statistical Association, 79,* 489–500.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* New York: W. H. Freeman.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science, 134,* 2011–2017.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs, 76*(Whole No. 547).

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29,* 261–265.

Paquette, L., & Kida, T. (1988). The effect of decision strategy and task complexity on decision performance. *Organizational Behavior and Human Decision Processes, 41,* 128–142.

Pauker, S. G., & Kassirer, J. P. (1987). Medical progress: Decision analysis. *New England Journal of Medicine, 316,* 250–258.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92,* 382–402.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 3,* 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). The adaptive decision maker: Effort and accuracy in choice. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 129–153). Chicago: University of Chicago Press.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72,* 346–357.

Pitz, G. F. (1974). Subjective probability distributions for imperfectly known quantities. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 29–41). Hillsdale, NJ: Erlbaum.

Pitz, G. F., & Sachs, N. J. (1984). Judgment and decision: Theory and application. *Annual Review of Psychology, 35,* 139–163.

Politser, P. (1981). Decision analysis and clinical judgment: A reevaluation. *Medical Decision Making, 1,* 363–389.

Politser, P. E. (1984). Explanations of statistical concepts: Can they penetrate the haze of Bayes? *Methods of Information in Medicine, 23,* 99–108.

Politser, P. E. (1987). Medical education for a changing future: New concepts for revising texts. *Medical Education, 21,* 320–333.

Politser, P., & Fineberg, H. W. (1987). *Toward predicting the value of a medical decision analysis.* Paper presented at the International Symposium of Forecasting, Boston.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty.* Reading, MA: Addison-Wesley.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 909–138.

Reynolds, J. H. (1987). *Computing in psychology: An introduction to programming methods and concepts.* Englewood Cliffs, NJ: Prentice-Hall.

Ronis, D., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40,* 192–218.

Rorer, L. G., & Dawes, R. M. (1982). A base rate bootstrap. *Journal of Consulting and Clinical Psychology, 50,* 419–425.

Rosen, A. (1954). A detection of suicidal patients: An example of some limitations in the prediction of infrequent events. *Journal of Consulting Psychology, 18,* 397–403.

Saks, M. J., & Kidd, R. F. (1980). Human information processing and adjudication: Trial by heuristics. *Law and Society Review, 15,* 123–160.

Sarbin, T. R. (1986). Prediction and clinical inference: Forty years later. *Journal of Personality Assessment, 50,* 362–369.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Savage, L. J. (1972). Diagnosis and the Bayesian viewpoint. In J. A. Jacquez (Ed.), *Computer diagnosis and diagnostic methods* (pp. 131–138). Springfield, IL: Charles C Thomas.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66,* 178–200.

Schank, R. C. (1984). *The cognitive computer: On language, learning, and artificial intelligence.* Reading, MA: Addison-Wesley.

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica, 68,* 203–215.

Shortliffe, E. H. (1986). Update on ONCOCIN: A chemotherapy advisor for clinical oncology. *Medical Informatics, 11,* 15–21.

Shortliffe, E. H. (1987). Computer programs to support clinical decision making. *Journal of the American Medical Association, 258,* 61–66.

Showers, J. L., & Chakrin, L. M. (1981). Reducing uncollectible revenue from residential telephone customers. *Interfaces, 11,* 21–31.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129–138.

Simon, H. A. (1979). Information processing models of cognition. *Annual Review of Psychology, 30,* 363–396.

Simon, H. A. (1986). Report of the research briefing panel on decision making and problem solving. *Research Briefings 1986, National Academy of Sciences.* Washington, DC: National Academy Press. (Reprinted in *Interfaces,* 1987, *17,* 11–31)

Sisson, J. C., Schoomaker, E. B., & Ross, J. C. (1976). Clinical decision analysis: The hazard of using additional data. *Journal of the American Medical Association, 236,* 1259–1263.

Slovic, P. (1972). Psychological study of human judgment: Implications for investment decision making. *Journal of Finance, 27,* 779–799.

Slovic, P. (1982). Toward understanding and improving decisions. In W. C. Howell & E. A. Fleishman (Eds.), *Human performance and productivity: Information processing and decision making* (Vol. 2, pp. 157–183). Hillsdale, NJ: Erlbaum.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39.

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

Steiner, I. D. (1972). *Group process and productivity.* New York: Academic Press.

Svenson, O. (1985). Cognitive strategies in a complex judgment task: Analysis of concurrent reports and judgments of cumulated risk over different exposure times. *Organizational Behavior and Human Decision Processes, 36,* 1–15.

Swets, J. A. (Ed.). (1964). *Signal detection and recognition by human observers: Contemporary readings.* New York: Wiley.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99,* 181–198.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285–1293.

Szucko, J. J., & Kleinmuntz, B. (1981). Clinical versus statistical lie detection. *American Psychologist, 36,* 488–496.

Szucko, J. J., & Kleinmuntz, B. (1985). Psychological methods of truth detection. In C. P. Ewing, Jr. (Ed.), *Psychology, psychiatry, and the law: A clinical and forensic handbook* (pp. 441–466). Sarasota, FL: Professional Resource Exchange.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401–409.

Thaler, R. H. (1986). The psychology and economic conference handbook: Comments on Simon, on Einhorn and Hogarth, and on Tversky and Kahneman. *The Journal of Business, 59*(4, Part 2), S279–S284.

Toda, M. (1962). The design of a fungus eater: A model of human behavior in an unsophisticated environment. *Behavioral Science, 1,* 164–183.

Toda, M. (1980). What happens at the moment of decision? Meta decisions, emotions, and volitions. In J. Sjoberg, T. Tyszka, & J. A. Wise (Eds.), *Human decision making* (Vol. 2). Bodafors, Sweden: Doxa.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch and by Hammond, Hursch, and Todd. *Psychological Review, 71,* 528–530.

Turner, D. R. (1966). Predictive efficiency as a function of amount of information and level of professional experience. *Journal of Projective Techniques and Personality Assessment, 30,* 4–11.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

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

Tversky, A., Sattah, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge University Press.

Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science, 29,* 151–173.

Waterman, D. A. (1986). *A guide to expert systems.* Reading, MA: Addison-Wesley.

Watson, C. G. (1967). Relationship of distortion to DAP diagnostic accuracy among psychologists at three levels of sophistication. *Journal of Consulting Psychology, 31,* 142–146.

Wertheimer, M. (1985). A gestalt perspective on computer simulations of cognitive processes. *Computers in Human Behavior, 1,* 19–33.

Wiehl, L. (1989, January 20). Computers assuming new notes at law firms. *New York Times,* p. 24.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.

Wiggins, N., & Hoffman, P. J. (1968). Three models of clinical judgment. *Journal of Abnormal Psychology, 73,* 70–77.

Wilkinson, L., Gimbel, B. R., & Koepke, D. (1982). Configural self-diagnosis. In N. Hirschberg & L. G. Humphreys (Eds.), *Multivariate applications in the social sciences* (pp. 103–113). Hillsdale, NJ: Erlbaum.

Winkler, R. L. (1986). Expert resolution. *Management Science, 32,* 298–303.

Winkler, R. L., & Makridakis, S. (1983). The combination of forecasts: Some empirical results. *Journal of the Royal Statistical Society,* Series A(146, Part 2), 150–157.

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 345–353.

Wright, G. (Ed.). (1985). *Behavioral decision making.* New York: Plenum Press.

Wright, G., & Ayton, P. (1986). Subjective confidence in forecasts: A response to Fischhoff and MacGregor. *Journal of Forecasting, 5,* 117–123.

Wrightsman, L. R. (1987). *Psychology and the legal system.* Monterey, CA: Brooks/Cole.

Yates, J. F., & Curley, S. P. (1985). Conditional distribution analyses of probabilistic forecasts. *Journal of Forecasting, 4,* 61–73.

Received September 28, 1988
Revision received March 16, 1989
Accepted June 6, 1989 ■