DECISION MAKING

Paul Slovic, Sarah Lichtenstein, and Baruch Fischhoff

Decision Research
A Branch of Perceptronics
1201 Oak Street
Eugene, Oregon 97401

In R. C. Atkinson, R. J. Herrnstein, G. Lindzey and R. D. Luce (Eds.), Stevens' Handbook of Experimental Psychology, second edition. New York: Wiley, in preparation.

December 1983

Acknowledgement

This research was supported by the Office of Naval Research under Contracts N014-80-C-0150 and N0014-82-C-0643 to Perceptronics, Inc.

Discussions with Ward Edwards, Ola Svenson, and Amos Tversky helped us greatly with organizing this chapter. Tversky, along with Duncan Luce, Don MacGregor, and Nils-Eric Sahlin, gave us very helpful comments on an early draft. Nancy Collins and Geri Hanson provided invaluable secretarial and technical assistance. We thank them all.

TABLE OF CONTENTS

INTRODUCTION

Although decision making has been studied for centuries by philosophers, mathematicians, economists, and statisticians, it has a relatively short history within experimental psychology. The first extensive review of the theory of decision making was published in the Psychological Bulletin by Edwards (1954). This paper introduced psychologists to the "exceedingly elaborate, mathematical and voluminous" (p. 380) economic literature on risky and riskless choice, utility, and game theory and reviewed the handful of relevant experimental studies then in existence.

Edwards' review was followed by a rapid proliferation of theories of choice and decision-making, along with carefully controlled experiments designed to test these theories. This work followed two parallel streams. One of these streams, the theory of riskless choice, had its origins in the notions of utility maximization put forth by Jeremy Bentham and James Mill. The first formal economic theories based on these notions assumed that decision makers are (a) completely informed about the possible courses of action and their consequences, (b) infinitely sensitive to differences in alternatives, and (c) rational in the sense that they can weakly order the possible courses of actions and make choices so as to maximize something--usually designated by the term utility. This stream of thought began to be interesting from the point of view of experimental psychology when theorists such as Thurstone (1927), Luce (1959), and Restle (1961) introduced modifications designed to capture the fact that decisions are characterized by inconsistency. Recognition of this fact has led to the development of stochastic theories of choice and a rich body of experiments designed to test those theories.

The second stream, the theory of risky choice, deals with decisions made in the face of uncertainty about the events that will determine the outcomes of one's actions. Maximization also plays a key role in these theories, but the quantity to be maximized becomes, due to the uncertainty involved, <u>expected utility.</u> Tests of the theory that individuals behave so as to maximize expected utility have been the topic of hundreds of studies, most of which studied reactions to well-defined manipulations of simple gambles as the basic experimental paradigm.

During the period between 1955-60, another development was taking place that was to have a profound influence on the study of decision making. This was the work of Simon (1956), who sharply criticized the notion of maximization as used in expected utility theory. Simon argued that actual decision making behavior is better described in terms of "bounded rationality." A boundedly rational decision maker attempts to attain some satisfactory, though not necessarily maximal, level of achievement, a goal that was labeled "satisficing." Simon's conceptualization highlighted the role of perception, cognition, and learning in decision making and directed researchers to examine the psychological processes by which decision problems are represented and information is used in action selection.

In recent years, the information-processing view has dominated the empirical study of decision making. Both streams of research, on risky and riskless choice, have been merged in a torrent of studies aimed at understanding the mental operations associated with judgment and decision making. The result has been a far more complicated portrayal of decision making than that provided by the utility maximization theory. It is now generally recognized that, although utility maximization can predict the outcomes of some decision-making processes, it provides only limited insight into how decisions are made. This descriptive limitation does not necessarily mean that utility maximization is not a valid principle for indicating how decisions should be made. Indeed, utility theory still forms the basis for the analysis of many applied decision problems. Increasingly, though, empirical evidence has prompted questioning of previously accepted tenets of rationality.

In sum, the theoretical status of the field of decision making is now undergoing a period of reexamination and criticism. Nevertheless, a coherent body of empirical findings exists and is beginning to be applied toward the solution of important practical problems faced by individuals, organizations, and societies, in the world outside of the laboratory. The path leading to this state of affairs is described in this chapter.

The chapter begins with a review of research describing the decision maker's subjective representation of the problem--the available alternatives, the possible outcomes of the decision, the environmental states that determine those outcomes, and the uncertainties surrounding those states and outcomes. Following this, we examine theories of the decision-making process, starting with models for deciding among simple, single-attribute alternatives and proceeding to models for handling more complex options. Some of these models are prescriptive, concerned with identifying courses of action that are logically consistent with the decision maker's expectations and values, whereas other models attempt to describe how people incorporate these expectations and values into their decisions. Some of these descriptive models closely resemble their normative counterparts. Others describe in detail the mental operations occurring during the decision-making process and thus look quite different. The chapter concludes with some speculations about a new view of preference emerging from empirical observations of decision making. This view poses a challenge to existing theories and to the decision-aiding technologies that have been derived from these theories.

COMPREHENDING THE DECISION ENVIRONMENT

This section presents research describing people's comprehension of the world in which a decision is being made. It focuses on predecisional activities. Perhaps the most important of these activities is problem structuring, in which the decision maker specifies (a) the possible actions, (b) the states of the world relevant to the decision, and (c) the outcomes contingent on both the chosen action and the states of the world that can occur. After structuring the problem, decision makers must consider the probabilities of the possible states of the world and the subjective values associated with the potential outcomes. In doing so, they must infer the causes, effects, and overall predictability of probabilistic phenomena. These inferences are in part generated by induction and in part deduced from experience. These various aspects of the decision environment are considered below.

Problem Structuring

A full description of a decision maker's intuitive problem structure would consist of all the options, consequences, and uncertainties considered in the course of reaching a decision. Although structuring is a key problem for decision makers, it has been less of a problem for decision researchers. In many experiments, subjects are presented with a predetermined problem, explicitly specifying all structural elements. This strategy allows researchers to focus on how people evaluate and integrate those components, at the expense of shedding light on how they identify them in less structured situations.

Gettys and his colleagues have undertaken an extensive research program on problem structuring. To look at hypothesis generation, Gettys, Manning, Mehle, and Fisher (1980) asked subjects to list all possible hypotheses about the cause of a problem such as an automobile malfunction. The subjects then estimated the probability that the actual cause of the problem was included in the list of generated hypotheses. Other tasks asked for hypotheses about students' undergraduate majors, workers' occupations, animals, and geographical areas. In all of these studies, subjects consistently generated hypothesis sets that lacked important hypotheses. At the same time, however, subjects regarded these impoverished sets as far more complete than they actually were. These general conclusions held for experts as well as lay subjects. Gettys et al. suggested using reference lists to facilitate the structuring process.

Not only do subjects fail to generate important hypotheses, they fail to recognize such omissions in lists generated by others. Fischhoff, Slovic, and Lichtenstein (1978) asked people to evaluate the completeness of an experimenter-generated hierarchical list of hypotheses (i.e., fault tree) describing why a car might fail to start (see Figure 1). Some subjects were given "pruned" trees in which half the major components were omitted. Subjects were asked to assess the probability that each of the presented components, including one called "all other problems," would prove to be the cause of car-starting failure. For both naive subjects and experienced car mechanics, what was out of sight was largely out of mind: the probability assigned to the "all other problems" branch was not much larger for severely pruned trees (e.g., those lacking "battery charge insufficient," "fuel system defective," and "other engine problems") than for the full tree.

Insert Figure 1 about here

Action generation has been found to be as impoverished as hypothesis formation. Pitz, Sachs, and Heerboth's (1980) subjects generated, on average, less than one third of the actions the experimenters judged worth considering in the context of solving a personal problem (e.g., dealing with an obnoxious roommate). Gettys, Manning, and Casey (1981) also gave subjects realistic dilemmas (e.g., solving a university's parking problem) and asked them to respond with possible actions. Performance was evaluated by constructing a group decision tree, combining all the acts suggested by the subjects into a hierarchical structure (as in Figure 1). The major ideas (e.g., increase available space for parking) formed the "limbs" of the tree and the variations of these ideas (e.g., build parking structures) formed the "branches" and "twigs" (e.g., build underground structures). It was found that individual subjects failed to generate many limbs and branches of the group decision tree, all of which they could, in principle, have thought of.

Gettys, Manning, and Casey also evaluated the quality of act-generation. A separate group of subjects estimated the utility of the acts generated by subjects in the act-generation experiment. These utility estimates were used to calculate the potential cost of failing to generate important limbs and branches. This analysis indicated that omissions were not only numerous, but also quite serious. In a follow-up study, Pliske, Gettys, Manning, and Casey (1982) found that providing monetary incentives for quantity and quality of actions produced did not improve performance.

Inability to conceive of important courses of actions has been found to degrade decision making in important settings outside of the laboratory. For example, field studies have shown that residents of flood plains are typically unaware of the range of actions that could be taken to reduce the risk of flood damage or insure against it (Kunreuther et al., 1978; White, 1964).

Uncertainty

Because many decisions involve uncertainty, a large literature exists on assessing and using probabilities (see, e.g., Kahneman, Slovic & Tversky, 1982). Two views of probability underlie this work. In the first, probability is defined as the limit of relative frequency. In the second, all probabilities are subjective. They are degrees of belief that are coherent; that is, they obey the probability laws. This view has been championed by de Finetti (1937/1980) and by Savage (1954), who incorporated it into utility theory. This "subjectivist," "personalistic," or "Bayesian" approach was introduced to psychologists by Edwards, Lindman and Savage (1963).

Although the battle between frequentists and subjectivists has raged for years among mathematicians and statisticians (e.g., de Finetti, 1980), its effect upon experimental psychology has been limited. Researchers have elicited probabilities for a variety of decision problems. Some of these allow an objective definition of probability, whereas others deal with unique events for which only subjective probabilities are meaningful. Edwards and his associates used a core principle in the subjectivist theory, Bayes' Theorem, as the basis for an optimal model of information processing (Slovic & Lichtenstein, 1971), but outside of that research tradition, it is rarely clear which meaning of probability is held by an experimenter. In most cases, the distinction is immaterial. Thus, for example, if you know that an urn contains 4 red balls and 6 white balls, both camps would agree that there is a probability of .4 that a ball blindly drawn from the urn will be red. One hint regarding the experimenter's view of probability is sometimes found: subjectivists are more likely to refer to the subjects assessing a probability (implying an internal search), while frequentists refer to subjects estimating the probability (implying the existence of an external answer).

Eliciting Probabilities

In everyday speech, people often use verbal phrases, such as "likely" or "improbable," to express their degree of uncertainty. Attempts to discover the numerical equivalents of these labels have shown great variability across people (Lichtenstein & Newman, 1967). The problem that this causes for communication, even among experienced forecasters, points to the need to use numbers or to develop consensual verbal labels (Beyth-Marom, 1982a).

If one is going to elicit numerical probabilities, a variety of approaches are possible. These include a simple probability scale between 0 and 1, an unlabeled rod with a moveable pointer, a logarithmically-spaced probability scale, verbally given odds, and a logarithmically-spaced odds scale (DuCharme & Donnell, 1973). There are also indirect methods such as response time (Geller & Pitz, 1968), choices among bets (Jensen & Peterson, 1973), or strength of a handgrip squeeze (Shapira, 1975). A moderate level of agreement has typically been found in studies comparing these direct and indirect methods (Wallsten & Budescu, 1983).

One theoretically based way to infer probabilities from bets is by means of proper scoring rules. These are functions that evaluate assessed probabilities according to both the degree of confidence and the outcome of the event being assessed. A scoring rule is "proper" if the only strategy for maximizing one's expected score is to state one's true belief (Shuford, Albert & Massengill, 1966). Proper scoring rules can be used to construct a list of payoffs for two-outcome bets. Each pair of payoffs is the pair of scores corresponding to a particular probability. The list is presented to subjects who are asked, for each event of interest, to pick the pair of scores they most prefer as outcomes for a bet, with the payoffs to be determined by the occurrence or non-occurrence of the event. If the heroic assumption is made that subjects choose in such a way as to maximize their expected earnings, then their subjective probabilities for events can be inferred from their choices (Jensen & Peterson, 1973).

An elicitation technique frequently used by professional decision analysts is a probability wheel (Spetzler & Staël von Holstein, 1975). This small device is a disc with two sectors of different colors, such that the relative proportion of the colors is easily changed. To assess the probability of event E, the wheel is adjusted until the assessor is indifferent between two (imaginary) bets, for example:

- Win $100 if E occurs; otherwise win nothing.
- Win $100 if a spinner, spun on the wheel, lands in the red sector; otherwise win nothing.

The probability inferred from this operation is equal to the proportion of red in the circle. This technique and others that similarly emphasize the relationship between probabilities and bets are the methods preferred by subjectivists. Their practical advantage has yet to be demonstrated.

The simplest method is to ask subjects to produce numerical probabilities after minimal instruction in the meaning of probability (e.g., "a response of .6 means that there's a 60% chance...a response of 1.0 means you're completely sure....") or no instruction at all ("what's the probability that....").

Surprisingly, amount of instruction appears to make little difference (Lichtenstein & Fischhoff, 1980).

Variants on these methods are available for expressing uncertainty about the value of an uncertain quantity (e.g., next year's interest rates, tomorrow's temperature). Ideally, subjects would draw subjective probability density or cumulative density functions. A less demanding approach, called the <u>fractile</u> method, has the experimenter specify several probabilities (e.g., .05, .25, .50, .75, and .95). For each probability, the subject then names a value of the uncertain quantity, such that the specified probability is the probability that the true value of the uncertain quantity is less than the subject's named value. Thus, for example, to elicit the .25 fractile for the population of the United States, the instructions might say "Write a population value such that there's a 25% chance that the true population is smaller than the one you write."

In the <u>fixed interval</u> method, the experimenter specifies several ranges of the uncertain quantity and asks the subject to assign a probability to each range. This method cannot easily be compared with the fractile method because the experimenter's choice of segments may provide information to the subject. For example, the following two partitions give quite different hints regarding the population of the United States:

___ less than 100 million	___ less than 1 million
___ 100 to 200 million	___ 1 to 10 million
___ 200 to 300 million	___ 10 to 100 million
___ 300 to 400 million	___ 100 million to 1 billion
___ more than 400 million	___ more than 1 billion

Evaluating Probabilities

A large number of experimental studies have examined how and how well people assess probabilities for tasks involving single and multiple events, uncertain quantities, and compound and conditional probabilities.

Frequency-based probabilities. Several studies (e.g., Robinson, 1964; Shuford, 1961) have found that people are quite good at estimating the relative proportions of binary events, displayed either sequentially (e.g., sequences of two rapidly flashing lights) or simultaneously (e.g., brief presentations of 400-element matrices of horizontal and vertical lines). To the extent that a probability is viewed as the limit of a relative frequency, these results suggest that within a reasonable range, say, .05 to .95, people are adept at assessing probabilities when the events are unambiguously presented in a short span of time. Estes (1976; Whitlow & Estes, 1979) has proposed a limited-capacity, multiple-trace model of memory to account for such findings.

When the relative frequencies are not presented directly, people must search their memories for instances. For such cases, Tversky and Kahneman (1973) have proposed that frequency and probability assessments are based on the ease by which instances come to mind. They called this process the "availability heuristic." In addition to being easy to apply, this heuristic is usually valid, because frequent events typically come more easily to mind. However, because availability is also affected by subtle factors unrelated to likelihood, such as familiarity, recency, and emotional salience, reliance on it may result in biased assessments. For example, Lichtenstein, Slovic, Fischhoff, Layman and Combs (1978) found that the frequencies of dramatic, well-publicized causes of death such as accidents, natural disasters, fires and homicides were overestimated and the frequencies of less-dramatic causes of death such as stroke, diabetes, emphysema, and asthma were underestimated.

Probability as confidence. Some events seem unique, or so nearly unique that it is difficult to conceptualize them as arising from a set of events with relative frequencies. For example, what is the probability that in 1997 the President of the United States will be a Republican? Probabilities for such events can be interpreted as degrees of belief or degrees of confidence. For such probabilities, there is no "right answer;" different people can justifiably hold different degrees of confidence in the same proposition. However, a kind of validity, called calibration, can be examined in a large collection of probability assessments. A set of probability assessments is said to be well calibrated if, in the long run, for all the events to which a probability of .XX was assigned, XX% of the events occur.

A large literature exists on calibration (for a review, see Lichtenstein, Fischhoff & Phillips, 1982). In a typical calibration experiment, subjects are given two-alternative general-knowledge questions (e.g., Which is longer, the Suez Canal or the Panama Canal? Are cabbage butterflies white or yellow?). For each question, they first select the alternative they believe to be the correct answer and then assess the probability that their chosen answer is, in fact, correct. These data are analyzed by computing the proportion correct for each probability value (or range of probabilities, such as .60-.69). The most common finding is that people's confidence in their knowledge is somewhat (but only somewhat) related to the accuracy of that knowledge. As confidence increases so does accuracy. However, the relationship is imperfect, with increases in confidence being accompanied by smaller increases in proportion of correct responses. In the typical study, this insensitivity leads to overconfidence. For all but guesses (p=.5), the proportions correct are notably smaller than the assessed probabilities (see Figure 2). This overconfidence is related to the difficulty of the items; it decreases as the difficulty of the items decreases until, for very easy items, underconfidence occurs.

Insert Figure 2 about here

The certainty response, p = 1.0, is particularly misused by subjects. Even when they are instructed that 1.0 means "you are absolutely sure...," 10% to 30% of the answers to which certainty is attached are incorrect ones. Concerned that this result reflected only an insensitivity of the response scale--subjects who use only one-digit probabilities may use the response of 1.0 to mean "more than .95"--Fischhoff, Slovic and Lichtenstein (1977) studied the calibration of subjects who responded with odds (e.g., 9:1, 10,000:1) rather than probabilities. This change did not eliminate the overconfidence observed with extreme responses. Subjects often used high odds (about one quarter of all responses were odds of 1,000:1 or greater), and were too often wrong. For example, when using odds of 100:1, subjects were right only 80% of the time; for odds of 10,000:1, they were 89% correct.

Overconfidence in probability assessments has been remarkably robust in the face of other manipulations as well (Fischhoff, 1982a; Lichtenstein et al., 1982). It has been found with both men and women and with people more or less expert in the content of the items (when task difficulty is controlled). Variations in instructions, response mode, and item content seem to make little difference. Efforts to eliminate overconfidence with monetary incentives, through training, or by requiring subjects to list reasons why they might be wrong have led to no improvement, improvement with limited generalization, and modest improvement, respectively.

The major exception to overconfidence is the performance of weather forecasters making probabilistic forecasts of precipitation (Murphy & Winkler, 1977). As a group, they are magnificently well calibrated. This superiority might be attributed to their years of experience in giving probabilistic forecasts, the homogeneous content area, and the unambiguous and rapid outcome feedback they receive.

Although calibration appears to be a universally desirable quality for probability assessments, that is not true when the set of items are not independent and the assessor receives no outcome feedback (Kadane & Lichtenstein, 1982). As an extreme (and artificial) example, suppose there are two urns, one containing 80% red balls, the other 20% red balls. One urn is chosen at random and one ball at a time is sampled from it (with replacement). You are not told which urn was chosen, nor the color of any of the sampled balls, yet you are asked, for each ball sampled, to assess the probability that the ball is red. Most people would find it appropriate to assess $p = .5$ for every ball, even though this string of .5 assessments will surely not be well calibrated, since, in the long run, the proportion of red balls will be either .8 or .2.

Why are people overconfident when assessing the extent of their own knowledge? Three possible explanations link this phenomenon to findings or theories in the domain of cognitive psychology: (a) Fischhoff et al. (1977) noted the tendency for people to believe that their memories are faithful (if faded) copies of their experiences, whereas evidence suggests that memory is a reconstructive process in which errors are sometimes incorporated as facts.

(b) Pitz (1974) suggested that in a series of inferences the uncertainty in the earlier stages may not be carried over into the later stages. (c) Koriat, Lichtenstein and Fischhoff (1980) emphasized the degree to which people search their memory only for confirming, not disconfirming, evidence concerning an initially favored answer.

Having arrived at a degree of confidence, how is that feeling translated into a numerical response? Ferrell and McGoey (1980) proposed a signal detection model (Luce & Krumhansl, Chapter _) for this process. They assumed that, in the absence of feedback about the difficulty of the items, people will not change the set of cutoff values that determine the translation of certainty feelings into probabilities. This model predicts overconfidence with hard items and underconfidence with easy items.

Calibration may also be studied in the assessment of probability density functions for uncertain quantities. Here, calibration refers to the correspondence between the fractiles of the distribution and the proportions of true values falling below or between the fractiles. For example, good calibration implies that, in the long run, just 25% of the true values should fall below the .25 fractiles, whereas 98% should fall between the .01 and .99 fractiles. Many experiments (reviewed by Lichtenstein et al., 1982) show that here, too, assessors are overconfident, in that they tend to report overly tight distributions. In a typical study, 30 to 40% of the true values of general-knowledge uncertain quantities (e.g., how many foreign cars were imported into the U.S. last year?) fall outside the .01 to .99 interval of the assessed distribution, rather than the appropriate 2%.

The three explanations offered for overconfidence apply here as well, as does an explanation based on the <u>anchoring and adjustment</u> heuristic (Slovic, 1972). Having decided there were, say, one million foreign cars imported, you take that initial estimate as an anchor and adjust it up and down to arrive at the higher and lower fractiles. These adjustments are typically insufficient, failing to account for the many ways the initial estimate could be in error.

Hindsight bias. Another form of overconfidence emerges in experimental studies of retrospective judgment. Studies by Fischhoff (1975; 1982b) have shown that reporting the outcome of an event increases the perceived likelihood of that outcome. Moreover, it does so in such a way that people underestimate the effect of outcome knowledge on their beliefs. As a result, people believe that they would have seen in foresight the relative inevitability of the reported outcome which, in fact, was only apparent in hindsight. Thus they exaggerate the predictability of reported outcomes. Slovic and Fischhoff (1977) showed similar effects in evaluations of scientific research; once people hear the results of an experiment, they tend to believe they "knew all along" what the findings would be. Apparently, outcome information is assimilated with whatever else is known about the event in question in a way that makes it impossible to retrieve the perspective once held in foresight. Hindsight bias seems to be as hard to reduce as other forms of overconfidence (Fischhoff, 1982a). Education or warnings have little effect. However, forcing people to think about how they could have explained the event that did not happen reduces the bias somewhat.

Judging probability by representativeness. When an uncertain event or sample is generated from a parent population by some process (such as randomly drawing a sample from a population), studies have shown that people judge its probability "by the degree to which it is: (i) similar in essential properties to its parent population; and (ii) reflects the salient feature of the process by which it is generated" (Kahneman & Tversky, 1972, p. 431). Kahneman and Tversky have labeled this strategy for assessing probabilities the <u>representativeness</u> heuristic.

People using the representativeness heuristic can be led astray either by attending to characteristics that are normatively irrelevant or by disregarding characteristics that are normatively important. As an example of the first type of error, people judging possible outcomes of tosses of a fair coin consider HTTHTH to be more likely than the outcome HHHTTT, because the lack of apparent order in the former seems more representative of a random process. They also find HTTHTH to be more likely than HHHHTH, because the latter does not represent the fairness of the coin (Kahneman & Tversky, 1972).

The second type of error is exemplified by people's disregard of sample size, a characteristic of a sample that has no parallel in the population. Thus, people consider that a large hospital (in which about 45 babies are born each day) will be as likely as a small hospital (in which about 15 babies are born each day) to experience a day on which more than 60% of the babies born are male.

<u>Combining Probabilities</u>

<u>Conservatism</u>. In the 1960's a much-researched topic was the question of how well people use the information from data to update the probability that a hypothesis is true. This research (reviewed by Slovic & Lichtenstein, 1971) was based on a strong normative model, Bayes' theorem. Given several mutually exclusive and exhaustive hypotheses, H_i, and a datum, D, Bayes' theorem states that:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_i P(D|H_i)P(H_i)}$$

$P(H_i|D)$ is the posterior probability that H_i is true, taking into account the new datum, D, as well as all previous data. $P(D|H_i)$ is the conditional probability that the datum D would be observed if hypothesis H_i were true. For a set of mutually exclusive and exhaustive hypotheses, H_i, the values of $P(D|H_i)$ represent the impact of the datum D on each of the hypotheses. The value $P(H_i)$ is the prior probability of hypothesis H_i. It, too, is a conditional probability, representing the probability of H_i conditional on all information available prior to the receipt of D. The denominator serves as a normalizing constant.

The following hypothetical experiment, similar to one actually performed by Phillips and Edwards (1966), illustrates the Bayesian paradigm. Subjects see two bookbags, one containing 70 red poker chips and 30 blue poker chips, the other containing 30 red chips and 70 blue chips. The experimenter flips a coin to choose one of the bags and then begins to draw chips from the chosen bag. After each chip is drawn, the subject assesses the probability that the predominantly red bag is the one being sampled. The optimal responses are computed from Bayes' theorem and are compared with the subjects' responses. The most frequently documented result is that subjects' assessments are conservative, in the sense that the optimal posterior probability of the most likely hypothesis is far larger than subjects' assessment. In the example just described, a sample of 8 red chips and 4 blue chips produces, from Bayes' theorem, a posterior probability of .97. Most subjects give an assessment between .7 and .8 (Edwards, 1968).

Early explanations attributed conservatism to subjects' (a) misunderstanding of the data-generating process and, thus, of the diagnostic impact of the data, $P(D\ H_i)$; (b) inability to aggregate the information received; (c) unwillingness to "use up" the bounded probability response scale, knowing that more data were forthcoming. Each of these explanations received some empirical support.

Later explanations have rejected the view that the normative model provides a good first approximation of a descriptive model, arguing that subjects rely on simple rules arrived at through "groping attempts to ease cognitive strain and to pull a number out of the air" (Slovic & Lichtenstein, 1971, p. 714). For example, when the data are presented sequentially, one such simple strategy is to revise one's estimate by a constant amount, upwards for confirming data or downwards for disconfirming. When simultaneous (aggregate) samples are presented, subjects appear to employ the representativeness heuristic, judging the similarity between the sample and the two possible populations (Kahneman & Tversky, 1972). The most notable feature of the populations is the ratio of red to blue chips in each; the representativeness heuristic, therefore, puts heavy emphasis on the ratio of red to blue chips in the sample. However, in such cases it is the difference, rather than the ratio, between the frequencies of the two colors in the sample which determines the posterior probability.

A variant of the bookbag-and-poker-chip task presents inconclusive evidence drawn from uncertain sources. For example, Gettys, Kelly and Peterson (1973) sampled a single chip (one of 4 possible colors) from a small container (one of 4 kinds of containers) that itself had been drawn from one of two bookbags and asked subjects to assess the probability of each bookbag given the color of the sampled chip. Fifteen of 25 subjects followed a "best-guess" strategy that ignores the probabilities of all but the most likely hypothesis. Such simplification is a typical response to multi-stage inference tasks, and one that supports Pitz's (1974) attribution of overconfidence to the failure to carry forward uncertainty from earlier stages. Schum (1980) gave a detailed discussion of the logical structure of more complex inference tasks.

The base-rate fallacy. Bayes' theorem combines base rate information (prior probabilities) with indicator information (conditional probabilities). Meehl and Rosen (1955) pointed out that clinical psychologists often disregard base-rate information when making predictions of rare events (like suicide), but only recently has the problem received experimental attention.

Kahneman and Tversky (1973) had subjects assess the likelihood that an individual described in a brief personality description was an engineer or a lawyer. The individual was allegedly sampled at random from a group consisting of 70 engineers and 30 lawyers, or from a group consisting of 30 engineers and 70 lawyers. The odds that any particular description belongs to an engineer rather than to a lawyer should be higher in the first condition, where there is a majority of engineers, than in the second condition, where there is a majority of lawyers. In violation of Bayes' theorem, subjects in the two conditions produced essentially the same probability judgments.

Kahneman and Tversky attributed this neglect of base-rate information to a reliance on representativeness, expressed here as the similarity of the descriptions to one's mental image of an engineer or a lawyer. Their subjects did use prior probabilities correctly, however, when they had no other information. In the absence of a personality sketch, they judged the probability that an unknown individual is an engineer to be .7 and .3, respectively, in the two base-rate conditions. This pattern of reliance on representativeness whenever possible has been bourne out in subsequent studies (reviewed by Bar-Hillel & Fischhoff, 1981).

Neglect of base-rate information has also emerged in studies of the "cab problem" by Kahneman and Tversky (1972) and others:

> Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green.
>
> A cab was involved in a hit-and-run accident at night.
>
> A witness later identified the cab as a Green cab.
>
> The court tested the witness' ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found that the witness was able to identify each color correctly about 80% of the time, but confused it with the other color about 20% of the time. What do you think are the chances that the errant cab was indeed Green, as the witness claimed?

Using Bayes' theorem, one finds that effect of the prior probabilities (.85, .15) slightly outweighs the effect of the conditional probabilities (.8, .2); the normatively correct answer is:

$$P(\text{Green}|\text{Data}) = \frac{.8(.15)}{.8(.15) + .2(.85)} = .41$$

However, in several studies (e.g., Bar-Hillel, 1980), the median and modal answer to the cab problem was 80%, showing a disregard of base rates.

Bar-Hillel (1980) offered the most encompassing explanation of the base-rate fallacy, proposing that "...subjects ignore base rate information, when they do, because they feel that it <u>should</u> be ignored--put plainly, because the base rates seem to them <u>irrelevant</u> to the judgment that they are making" (p. 216). According to Bar-Hillel, apparent relevance can be produced not only by saliency, as in the engineer/lawyer problem above, but also by causal links among the data. For example, the cab problem can be modified to create a causal link with the base rate by replacing the sentence "Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green" with the sentence "Although the two companies are roughly equal in size, 85% of cab accidents involve Green cabs and 15% involve Blue cabs" (Tversky & Kahneman, 1982). The problem now suggests that Green drivers are more reckless or incompetent than Blue drivers. Although answers to this version were still highly variable, the median response was 60%, indicating that the base rate was less often ignored.

Conjunction problems. One of the simplest and most basic laws of probability is the conjunction rule: The probability of a conjunction, $P(A \cap B)$, cannot exceed the probabilities of its constituents, $P(A)$ and $P(B)$. Tversky and Kahneman (in press) have shown that people's intuitive judgments of probability violate the conjunction rule when the conjunction is more representative of the underlying process than is one of the two constituent events. For example, most subjects violated the conjunction rule in the following problem.

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Which of the following statements is more likely?
>
> • Linda is a bank teller
>
> • Linda is a bank teller and is active in the feminist movement.

The first of these answers seemed to fit the description of Linda so poorly that it was deemed less likely than the conjunction in the second answer, which added a detail fitting her description. Tversky and Kahneman argued that it is hard to advance any normative theory of inference that would view such behavior as acceptable. Earlier studies (e.g., Bar-Hillel, 1973; Beyth-Marom, 1981) have shown, either using direct estimates or choices among bets, that the probabilities of conjunctions (i.e., the probability that all of several uncertain events will occur) are overestimated and the probabilities of disjunctions (at least one event will occur) are underestimated. These violations, too, were traced to reliance on the representativeness heuristic.

Debiasing

A number of studies have attempted to eliminate judgmental biases in probability assessment. There are two possible goals for such studies: (a) to find practical ways to improve performance and (b) to test the robustness of the bias and thereby reveal something about the processes that produced it. These debiasing manipulations can be categorized according to whether they attribute the source of the bias to the task (for failing to elicit the subjects' extant knowledge), to the subjects (for lacking the requisite skills), or to a mismatch between subject and task (meaning that the subject cannot manage the task as presented, but could do better with a restructured version). These categories can be subdivided further into the set of debiasing categories described in Table 1.

Insert Table 1 about here

The list in Table 1 provides a way to generate experiments as well as a way to categorize debiasing studies that have already been conducted. A review of all published studies attempting to eliminate two biases (overconfidence and hindsight) revealed some consistent patterns (Fischhoff, 1982a). Manipulations that treat the bias as an experimental artifact have had little effect. Thus, for example, it does not help to raise the stakes, exhort subjects to work harder, or rework the instructions (providing they were already clear and fair). Nor is there any consistent improvement when the stimuli come from people's areas of expertise. Thus, although substantive expertise gives people many answers and tools with which to seek answers, it is not clear that it improves their judgment per se. On this important question, further research is needed.

It is, however, possible to improve performance somewhat through training that includes personalized feedback. Merely warning subjects about the problems that others experienced makes no difference (unless it is so directive as to tell subjects what to do). Finally, there has been some success with manipulations based on theories of the cognitive processes leading to the biases. For example, the belief that subjects are over-confident because they naturally tend to think of reasons justifying their answers led to a manipulation that required them to list reasons why they might be wrong (Koriat et al., 1980). In a related vein, Lopes (1982b) reported success in debiasing judgments in a Bayesian inference task by (a) analyzing the procedures used by untutored subjects, (b) warning subjects about the procedures they use that are inappropriate, and (c) providing information about appropriate procedures.

In debiasing research, an important distinction is between producing better judgments and producing better judges. One could produce better probability judgments on a calibration task with items of moderate difficulty by telling subjects to lower their responses. Unfortunately, that same advice would increase the underconfidence found with tasks having easy items. By contrast, training or looking for contradictory reasons offers some hope of improving the judgment process and being broadly useful.

Relatedness

Correlations

A key intellectual skill, indeed, the basis for a causal understanding of the world, is the ability to assess the interrelatedness of two variables. Jennings, Amabile and Ross (1982) distinguished between two versions of this skill. The first, data-based assessment, concerns the ability to detect covariation in novel data, about which one has no expectations. The second, theory-based assessment, concerns covariation estimates that are derived from one's a priori expectations or theories, rather than from any immediately available data.

Data-based assessment. In order to avoid the influence of prior expectations, studies of relatedness have typically presented subjects with data concerning two hypothetical dichotomous variables, such as could be represented in a four-fold table. The correct interpretation of correlation involves all four cells. When the cells are labeled a, b, c, d in the conventional way, the correlation is the difference between two conditional probabilities:

$$\frac{a}{a+c} - \frac{b}{b+d}$$

Oft-cited studies have indicated that subjects base their judgments of relatedness on only one cell (the "yes/yes" cell; Smedslund, 1963; Jenkins & Ward, 1965) or only two cells (the "yes/yes" and "no/no" cells; Ward &

Jenkins, 1965). Beyth-Marom (1982b) pointed out that the instructions to the subjects in the earlier studies may have accounted for these results. For example, Smedslund's instructions included the statement, "You are to concentrate entirely on symptom A and diagnosis F," which could be interpreted as telling the subjects to focus on only one cell, which most did. Ward and Jenkins' instructions included the sentence "Complete control means that whenever you seed, it rains, and whenever you don't seed, it doesn't rain," thus focusing on the a and d cells.

Using more neutral instructions, Shaklee and Tucker (1980) devised a series of four-fold tables that enabled them to infer the rule subjects were using. Subjects' accuracy in judging whether the variables were correlated was improved by instruction in the concept of covariation and by use of a response format that highlighted the conditional probabilities of events. Beyth-Marom (1982b) presented subjects with a list of explicit rules and asked subjects to choose the one that best fit their interpretation of a statement such as, "A paper published in a major biological journal reported that for one species of widely-distributed animals a strong relationship was found between the animal's skin color and the mean temperature in its territory". Beyth-Marom's study, like that of Shaklee and Tucker, found that subjects exhibited or chose rules involving all four cells, either the correct rule or a rule contrasting the sums of the two diagonals. Beyth-Marom further showed that the simpler incorrect rules were chosen more often with asymmetric variables, for which the name of one pole of the variable is the name of the variable (e.g., a disease is present or absent), than with symmetric variables (e.g., skin color is dark or light).

Multiple-cue probability learning (MCPL) studies have examined people's understanding of interconnectedness in tasks that involve continuous or many-valued variables in an environment that provides an opportunity for learning. In the typical study, subjects predict a numerical criterion on the basis of sets of numerical cues. The cues are related to the criterion probabilistically by adding an error term to the functional relationship between the cues and the criterion. Blocks of training trials, during which the criterion number is shown to the subject after each prediction, are alternated with test blocks, during which the subject makes predictions without receiving feedback. Results of MCPL studies show that: (a) subjects can learn to use linear cues appropriately; (b) learning of nonlinear functions is slow, especially when subjects are not forewarned that relations may be nonlinear; (c) subjects are inconsistent, particularly when task predictability is low; and (d) subjects fail to take proper account of cue intercorrelations. For a review of MCPL studies, see Brehmer (1979).

Theory-based assessment. The distinction between data-based and theory-based assessment is not always a firm one. Brehmer (1980a) showed that subjects in MCPL tasks are guided by specific hypotheses about the functional rule relating cues and criterion. These hypotheses appear to be sampled from a set based on previous experience and dominated by the positive linear rule. Testing of these rules shows inadequate appreciation of the probabilistic nature of the task. Subjects keep searching for deterministic rules that will account for the randomness in the task. Because there are none, they change rules frequently (i.e., become inconsistent) and eventually resample rules they had previously discarded.

The MCPL studies typically use artifical tasks, for which people have no firm expectations regarding the kinds of results that might be found. Chapman and Chapman (1967) hypothesized that, when people do have strong expectations, they often tend to find confirmation of their beliefs in settings that show quite the opposite. To demonstrate this, they created a set of 45 drawings from the Draw-A-Person test, a common diagnostic tool in clinical psychology. Each drawing was labeled with two symptom statements (e.g., "He is suspicious of other people") in such a way that each symptom statement appeared once with each drawing, producing zero correlation between any drawing characteristic and any symptom. The statements were collected from practicing clinicians regarding characteristics of drawings that they believed to be associated with the symptoms (e.g., "eyes atypical"). When the subjects, students in an introductory psychology class, listed characteristics that were associated with each symptom, they reported the same symptom-characteristic correlations as had the clinical psychologists. For example, the majority of both clinicians and naive subjects reported atypical eyes being associated with suspiciousness. This phenomenon, which the Chapmans labeled "illusory correlation," persisted even when the subjects were given more study time, were rewarded for accuracy, or were shown negative correlations between symptoms and diagnoses.

The shaping effect of preconceptions on the processing of subsequent data has been shown in a variety of other settings (Nisbett & Ross, 1980). It stands in sharp contrast to the neglect of base rates discussed earlier. The key difference between those two settings seems to be whether subjects create those prior beliefs themselves or receive them in the experiments.

Once people believe in the existence of a relationship, they may use it in prediction tasks. When the relationship is between continuous variables, statistical theory requires that the prediction be regressed to the mean to the extent that the relationship is not perfect. In practice, though, such regression is often absent. Subjects predicting variable Y from variable X tend to make their prediction of Y as extreme as that of X. Thus, for example, Kahneman and Tversky (1973) found that two groups of subjects gave virtually identical predictions of 10 students' GPAs (Grade Point Averages), although one group was given each student's percentile score on GPA (which is perfectly correlated with GPA), while the other group was given percentile scores of a "test of mental concentration" that was described as quite unreliable: "...when tested repeatedly, the same person could obtain quite different scores, depending on the amount of sleep he had the night before or how well he felt that day." A third group, given percentile scores for a "test of sense of humor," showed only a slight regression in the prediction of GPA. Tversky and Kahneman (1974) linked non-regressive prediction to the representativeness heuristic, which leads people to predict outcomes that are maximally representative of (similar to) the input.

Failure to appreciate the workings of regression effects is not limited to naive subjects working on unfamiliar tasks in psychological experiments. A number of articles have taken statistically trained researchers to task for failing to recognize possible regression artifacts in their data (e.g., Campbell & Boruch, 1975; Furby, 1973).

<u>The Lens Model</u>

The philosophy of <u>probabilistic functionalism</u> put forth by Brunswik (1952, 1956) has played an important role in conceptualizing how people attempt to understand and cope with uncertainty in the decision environment. Brunswik's main interest was the adaptive interrelationship between the person (judge or decision maker) and the decision situation. He developed the <u>lens model</u> to study simultaneously the uncertain, interdependent structure of the world and the judgments that people make about that structure.

The lens model, shown in Figure 3, gets its name from the symmetry between the environmental system and the organismic system. In the center are the cues: the proximal variables available as information for the judge. The cues are related to the criterion variable in the enviromental half of the lens and to judgments in the other half of the lens.

Insert Figure 3 about here

The first step in applying the lens model is to develop sets of stimuli that match, in their structure, the structure of the environment. For example, in such a representative design, cues that are interrelated in the world are similarly correlated across the stimulus set. These stimuli are presented to subjects whose task is to make wholistic judgments (estimates) of the criterion value for each stimulus profile of cue values. These judgments are then analyzed in terms of a model; the linear additive model is most commonly used. The analysis produces statistical measures of the key concepts: achievement (r_a, the correlation between the judgments and criterion values), ecological validity ($r_{e,i}$, the correlation between the cue values and the criterion), and cue utilization ($r_{s,i}$, the correlation between the cue values and the judgments). Further measures based on multiple regression analysis have been developed by Hursch, Hammond and Hursch (1964), Tucker (1964), and Dudycha and Naylor (1966).

The relationships studied within the lens model can be either data-based (as with the MCPL studies described above) or theory-based. Numerous studies have shown that the lens model in its simplest, additive linear, form is remarkably sucessful in fitting such diverse judgments as psychiatric and medical diagnoses, judicial decisions in workers' compensation and civil liberties cases, and roll call votes of U.S. Senators, as well as judgmental evaluations of job performance, graduate school applicants, suicide risk, financial status, stock prices, theatrical plays, and trout streams (Slovic & Lichtenstein, 1971; Slovic, Fischhoff & Lichtenstein, 1977). The variety of ways in which the model can be used to understand and improve judgments has been summarized by Hammond, Stewart, Brehmer, and Steinmann (1975).

Randomness

An understanding of probabilistic relationships requires an understanding of the lack of a relationship or randomness. When people judge the randomness in a sequence of dichotomous events or attempt to create random sequences, they demonstrate a consistent bias in believing that random sequences have more alternations than they actually do (Evans & Pollard, 1982; Wagenaar, 1972).

A related expression of these same intuitions is the gambler's fallacy: the belief that following a period in which one possible outcome of a random sequence occurs less often than its expected frequency, its future occurrence is more likely. Also known as the "negative recency effect" (Jarvik, 1951), this belief has been frequently reported: "Of course, after the red has come up ten times in a row, hardly anyone will persist in betting on it" (Dostoevsky, 1866/1964, p. 146).

Lopes (1982a) pointed out some problems in the design and interpretation of psychological studies of subjective randomness. She criticized the relatively narrow conception of randomness that underlies these experimental studies, and contrasted this with the conceptions found in philosophical and mathematical treatments.

The Place of Bias in Judgment

Defining optimality. To claim that responses are biased requires that optimality be well defined, with a normative theory of how probabilistic judgments should be made and a substantive theory specifying how a particular problem is interpreted in terms of the normative theory. In many studies the rules of Bayesian inference have been used to define optimality. The normative status of Bayesian inference is not, however, unquestioned. Some philosophers (e.g., Cohen, 1981) have offered alternative formulations and even those who accept the Bayesian framework (e.g., Diaconis & Zabel, 1982) acknowledge certain limitations. An emerging compromise position (Shafer & Tversky, in press) is that a normative theory should not be treated as an absolute standard, valid in all situations. Rather, there are alternative theories, each suited for particular situations, with the choice between them determined on grounds of practicality.

More serious questions have been raised about whether specific instances of observed behavior can be properly described as suboptimal, under the assumption that the decision maker is attempting to follow an optimal model. In order to make an unambiguous interpretation of an act, one needs to know how the decision maker construed the model for that situation. Inferences that appear to be biased may prove to be quite legitimate given a better understanding of how the decision maker's hypotheses were formulated or of what actions hinged on those inferences. Moreover, a full account of suboptimal judgment must consider each component of the model, lest a difficulty with one be misattributed to another.

Fischhoff and Beyth-Marom (1983) examined the components of Bayesian inference (e.g., hypothesis formation, assessing prior odds and likelihood ratios), calling into question a number of attributions of judgmental bias. In some cases, the bias was other than had been claimed, whereas in others there may have been no bias at all. They also found cases in which apparently diverse effects proved to be special cases of a single judgmental bias. The most powerful of these "metabiases" is the tendency to ignore P(D H) when evaluating evidence. Finally, they suggested that the "confirmation bias" (Doherty, Mynatt, Tweney & Schiavo, 1979; Snyder & Swann, 1978; Wason, 1960, 1968) should not be viewed as a single bias towards seeking confirmation of an hypothesis but as several different patterns of information search and evaluation group under a common label.

<u>Learning from experience</u>. In probabilistic tasks for which optimal responses have been satisfactorily defined, research has shown that people often lack or fail to exhibit the cognitive skills required for optimal performance. Moreover, people do not seem to realize their inabilities, showing what Dawes (1976) called "cognitive conceit." In part, these failings may be traced to a lack of formal schooling in probabilistic inference (Beyth-Marom, in press). However, Brehmer (1980b) and Einhorn and Hogarth (1978) have attributed cognitive conceit to difficulties inherent in learning from an uncertain environment.

First, learning requires the formation of concepts or hypotheses about what is to be learned. In the laboratory such concepts are made evident by the experimenter. In the real world, instances are multidimensional and the concepts are not manifest. Concepts about uncertainty may be particularly slow to form. We could not function in our environment without searching for regularities and finding deterministic rules. The attention given to regularity may cause us to overlook or even deny the existence of unpredictabilty.

Once generated, hypotheses must be evaluated. Even in laboratory studies, hypothesis evaluation has been found flawed (Evans, 1982; Fischhoff & Beyth-Marom, 1983). When deterministic prediction rules are used in uncertain environments, an inaccurate prediction may be interpreted as evidence that the rule is wrong, rather than as an acceptable and expected error in a world that is only partially predictable. When probability assessment is flawed, a few surprising outcomes cannot pinpoint the error. Even in ideal circumstances, an evaluation of the hypothesis of overconfidence in probability assessment requires a great amount of data.

The feedback necessary for learning may be unavailable or distorted. Some outcomes are never known. Others are so delayed that they have little impact. Whenever memory is required to compare outcomes against predictions, inaccuracies may arise. For example, memory of one's prediction may be biased by knowledge of subsequent events (Fischhoff, 1975). Moreover, outcomes may be stored in memory according to their content rather than as examples of inferential rules, making their recall difficult when evaluating one's predictive skills (Tversky & Kahneman, 1974).

Further complications arise when judgment leads to action, such as selecting applicants for a job. Einhorn and Hogarth (1978) have shown how such choices bias the feedback received, as a function of four variables: (a) judgmental validity (how good the judgments are), (b) selection ratio (what proportion of cases are selected), (c) base rate (what proportion of cases meet the criterion), and (d) treatment effects (what effect selection has on the success rates for selected cases).

When the base rate of success is high, one will, of course, observe a high proportion of successes (positive hit rate), almost regardless of the selection ratio or the validity of the judgments. When the selection ratio is low (few applicants are accepted), the improvement in proportion of successes is a steep function of judgment validity. Thus, for example, if the selection ratio is .1, the base rate is .5, and judgmental ability is zero (judgments uncorrelated with the success criterion), of those accepted, only 50% will succeed. But if the judgments correlate .4 with the criterion, over 75% of the selections will succeed, and if judgmental validity is .6 (i.e., only 36% of the variance associated with the success variable is explained by our judgments), 90% of the selections will succeed.

The positive hit rate is biased further when the act of selection improves the selected object or individual, as when new employees are trained to make them more capable at the job or when research grants improve both the quality and the quantity of ther recipients' research. Einhorn and Hogarth showed that the effect of such enhancement is largest when the positive hit rate is otherwise low.

In sum, the world does not work in a way that helps us recognize our deficiencies in probabilistic judgment. As a result, we maintain an exaggerated sense of our judgmental prowess.

THEORIES OF DECISION MAKING

Single Attribute Risky Models

Decision problems under conditions of risk can be conveniently represented by means of decision matrices and decision trees. In the decision matrix, the rows correspond to alternative acts that the decision maker can select and the columns correspond to possible states of nature. The cells of the matrix describe the consequences contingent upon the joint ocurrence of a decision and a state of nature. A simple illustration for a traveler is given in Table 2. An analogous pictorial representation takes the form of a decision tree (see Figure 4). Trees have the advantage of being better able to represent complex problems involving sequences of decisions over time (Figure 5).

Insert Table 2 and Figures 4 and 5 about here

Early Theories

Because it is not always possible to make a decision that will turn out best in any eventuality, decision theorists view choice alternatives as gambles and try to determine the best bet. With the development of probability theory during the 17th century, the best bet came to be defined as the alternative that maximizes the expected value of the decision. That is, it maximizes the quantity

$$EV(A) = \sum_{i=1}^{n} P(E_i)V(x_i)$$

where EV(A) represents the expected value of a course of action that has consequences $x_1, \ldots x_i, \ldots x_n$ depending on events $E_1, \ldots E_i, \ldots E_n$, $P(E_i)$ represents the probability of the i-th outcome of that action, and $V(x_i)$

represents the stated value of that outcome (for example, the monetary gain or loss). The expected value of a gamble can be viewed as the average outcome resulting from playing it a large number of times. Gambles are often labeled as favorable, unfavorable, or fair depending on whether their expected values are positive, negative, or zero.

A little reflection shows that maximization of expected value is an inadequate model for describing people's behavior. People gamble in casinos even though the expected value of playing the games there is less than that of not playing. They buy insurance even though the premium costs more than the expected value of the undesirable risk that the insurance covers. Nor are people indifferent in their evaluation of fair bets. For example, most would reject the opportunity to toss a fair coin offering the possibility of either winning or losing $100.

Related observations led mathematician Daniel Bernoulli (1738/1954) to propose that people's actions are governed by the expected <u>utility</u>, rather than the expected value, of the gamble. Expected utility is determined by substituting the subjective worth or utility of each outcome in place of the values $V(x_i)$. Bernoulli went on to propose a specific logarithmic function to represent his notion that the more money one has, the less each additional increment is valued.

Given the selection of an appropriate utility function, substitution of utilities for actual monetary values can account for gambling and insurance decisions that fail to maximize expected value. This does not mean that people gamble or insure their property in order to maximize some utility function, only that their choices are consistent with utility maximization.

Modern Utility Theory

A major deficiency of Bernoulli's principle is that it provides no normative justification for maximizing expected utility on a single choice as well as on a long-run series of decisions. Modern utility theory, as formulated by von Neumann and Morgenstern (1947), provides such a justification. Von Neumann and Morgenstern showed that if people's preferences among gambles satisfied certain axioms, then their behavior could be described as the maximization of expected utility. Because the axioms embody basic principles of rational behavior, they provide a normative basis for the expected utility principle.

We shall present the von Neumann and Morgenstern axioms in the form given by Coombs, Dawes, and Tversky (1970). The axioms are formulated in terms of a preference-or-indifference relation, denoted $\succsim$, defined on a set of outcomes, denoted A. This set is enriched to include gambles, or probability mixtures, of the form (x, p, y), where outcome x is obtained with probability p and outcome y is obtained with probability 1-p. Given the primitives $\succsim$ and A, the following axioms are assumed to hold for all outcomes x, y, z in A and for all probabilities p, q that are different from zero or 1.

Axiom 1: (x,p,y) is in A.

Axiom 2: $\succsim$ is a weak ordering of A, where $\succ$ denotes strict preference, $\sim$ denotes indifference, and for all outcomes x,y,z the relation satisfies:

i. Reflexivity: $x \succsim x$.

ii. Connectivity: Either $x \succsim y$ or $y \succsim x$ or both.

iii. Transitivity: $x \succsim y$ and $y \succsim z$ imply $x \succsim z$.

Axiom 3: $[(x,p,y),q,y] \sim (x,pq,y)$.

Axiom 4: If $x \sim y$, then $(x,p,z) \sim (y,p,z)$.

Axiom 5: If $x \succ y$, then $x \succ (x,p,y) \succ y$.

Axiom 6: If $x \succ y \succ z$, then there exists a probability p such that $y \sim (x,p,z)$.

Utility theory can be viewed either as a normative theory, justified by the appeal of its axioms as ways in which one should make decisions, or as a descriptive theory. Hence the axioms can be examined from both perspectives.

The first axiom is what is technically called a closure property. It asserts that if x and y are available alternatives, so are all the gambles of the form (x,p,y) that can be formed with x and y as outcomes. Because gambles are defined in terms of their outcomes and their probabilities, it is assumed implicitly that $(x,p,y) = (y,1-p,x)$. The second axiom requires the preference-or-indifference relation to be reflexive, connective, and transitive. Reflexivity and connectivity are empirically trivial. Although transitivity is systematically violated in certain contexts, it is, nevertheless, a very compelling normative principle and a plausible descriptive hypothesis.[1]

Axiom 3 is a reducibility condition. It requires that the gamble (x,pq,y), in which x is obtained with probability pq and y with probability 1-pq, be equivalent, with respect to preference, to the compound gamble [(x,p,y),q,y], in which (x,p,y) is obtained with probability q and y with probability 1-q. Axiom 3 asserts, in effect, that the preferences depend only on the final outcomes and probabilities and not on the process by which the outcomes are obtained.

Axiom 4 is a substitutability condition. It states that if x and y are equivalent, then they can be substituted for one another in any gamble.

The fifth axiom asserts that if x is preferred to y, then it must be preferred to any probability mixture of x and y, which, in turn, must be preferred to y.

The sixth axiom embodies a continuity or a solvability property. It asserts that if y is between x and z in the preference order (i.e., $x \succ y \succ z$) then there exists a probability p such that the gamble (x,p,z) is equivalent to y. This axiom excludes the possibility that one alternative is "infinitely better" than another one.

Axiom 6 captures the relationships between probabilities and values and the form in which they compensate for each other. This form becomes transparent in the following theorem of von Neumann and Morgenstern.

THEOREM

If axioms 1-6 are satisfied, then there exists a real-valued utility function u defined on A, such that

1. $x \succsim y$ if and only if $u(x) \geq u(y)$.
2. $u(x,p,y) = pu(x) + (1-p)u(y)$.

Furthermore, u is an interval scale; that is, if v is any other function satisfying 1 and 2, then there exist numbers $a>0$ and b such that $v(x) = au(x)+b$.

Thus the theorem guarantees that whenever the axioms hold, there exists a utility function that (a) preserves the preference order and (b) satisfies the expectation principle: the utility of a gamble equals the expected utility of its outcomes. Moreover, this utility scale is uniquely determined except for an arbitrary origin and unit of measurement.

The main contribution of modern utility theory to the analysis of decision making under risk is in providing a justification for the Bernoullian expected utility principle. This justification does not depend on long-run considerations; it is applicable to unique choice situations. Furthermore, the axiomatic structure highlights those aspects of the theory that are critical for both normative and descriptive applications.

Subjective expected utility. One limitation of the von Neumann and Morgenstern theory is its treatment of probability. Although Ramsey (1931) and de Finetti (1937/1980) had laid the groundwork for a subjective theory based on the concept of probability as a degree of belief about the likelihood of an event, von Neumann and Morgenstern assumed the existence of known numerical probabilities for all events. A major advance occurred when Savage (1954) developed an axiomatic theory allowing simultaneous measurement of utility and subjective probability. By means of several powerful axioms, Savage proved the existence of a unique subjective probability function S, which obeys all the usual laws of probability, and an interval scale utility function u such that

1. $x \succsim y$ if and only if $u(x) \geq u(y)$ and
2. $u(x,E,y) = S(E)u(x) + [1-S(E)]u(y)$

where (x,E,y) denotes the gamble where x is obtained if E occurs and y otherwise. Edwards (1955) labeled this the subjective expected utility (SEU) model. Excellent discussions and critiques of Savage's theory are provided by Krantz, Luce, Suppes, and Tversky (1971) and Fishburn (1982a).

Risk aversion. Although the utility function under either the von Neumann-Morgenstern formulation or the Savage formulation can take on any shape, utility functions are often characterized as risk averse, risk prone, or risk neutral. A risk-averse function is one such that, for probability p,

$$U(px) > pU(x).$$

A risk prone function reverses this preference,

$$pU(x) > U(px),$$

whereas risk neutrality means indifference between the two quantities. A risk-averse utility function is concave; a risk-prone function is convex, and a risk-neutral function is linear with value. There is nothing in the theory to bar a person's utility function from being risk prone in some regions and risk averse in others.

A risk-averse person, that is, a person whose utility function is concave in the region of interest, will prefer, for example, to receive \$50 for sure over a 50-50 chance of receiving either \$100 or nothing; a risk-prone person will prefer the gamble and the risk-neutral person will be indifferent between the two options.

Measurement methods. Farquhar (1982) has discussed more than two dozen techniques for measuring utility based on variations of a few general methods. The standard gamble contrasts a bet (x,p,y) with a sure outcome, w, called the certainty equivalent. One of the four elements, w, x, y, or p, is omitted and the decision maker is asked to specify it so as to create indifference between the bet and the sure outcome. For example, consider a bet paying either \$0 or \$10 and a sure outcome of \$5. What probability of winning \$10 would make you indifferent between playing the bet and receiving \$5? A bit of algebra shows that if u(\$0) = 0 and u(\$10) = 1, the answer, p, to this question is equal to the utility of the sure outcome. For example, if you answered .8, then u(\$5) = .8 (thus exhibiting risk aversion). Because of this directness, standard gambles are often constructed with x being the best possible outcome and y being the worst possible outcome. The assessor is asked to supply p values for enough w values to reveal the utility curve, scaled from 0 to 1, to any desired degree of precision.

Alternatively, a series of bets paired with sure outcomes is presented and the decision maker indicates, for each pairing, preference for the bet, preference for the sure outcome, or indifference. Each such comparison provides a constraint that the utility function must satisfy. With a well-chosen set, sufficiently tight bounds on the utility function can be found.

Finally, all the above variations can be used with pairs of gambles rather than pairs consisting of a gamble and a sure outcome.

With the use of computers, multiple methods can be employed, in which sequences of comparisons are used to generate systems of equations whose solutions provide the utility function (Novick, Dekeyrel & Chuang, 1981). Built-in consistency checks show the decision maker the implications of previous choices and allow adjustments to be made to achieve consistency.

Tests of Utility Theory

Following the development of an axiomatic justification leading to measurable probability and utility functions, expected utility maximization gained great popularity, not only as a model of how people should behave, but also as a psychological or descriptive theory about how they actually do behave. The model's normative and descriptive properties have, over the years, steadily infiltrated theories in such diverse disciplines as economics, philosophy, finance, psychology, political science, and management science and have formed the basis of a new discipline, decision analysis, designed to help people make optimal decisions in situations of risk. Coombs (1975) observed that the fundamental role played by utility maximization demands that the theory be "probed relentlessly" (p. 65) and so it has been.

<u>Predictive tests</u>. Early empirical studies based on the von Neumann-Morgenstern and Savage formulations attempted to determine whether people's decisions were consistent enough so that utility functions could actually be constructed and used to predict other decisions. The first such study was conducted by Mosteller and Nogee (1951), who presented subjects with gambles constructed from possible hands of a poker game played with dice. If subjects rejected a gamble, no money changed hands; if a gamble was accepted, subjects won x if they beat the hand but lost a nickel otherwise. The actuarial probability of beating each hand was made available to the subjects. By varying the payoffs, the experimenter determined the value of x for which the subject was indifferent between the two alternatives. These indifference amounts were used to construct utility functions, with u(0) set equal to 0 and u (lose 5¢) set equal to -1. These utility functions predicted subjects' choices among new bets more accurately than did predictions based on the maximization of expected monetary value.

Whereas Mosteller and Nogee assumed that their subjects used the stated actuarial probabilities, Davidson, Suppes, and Siegel (1957) attempted to measure both utility and subjective probability. Their approach followed Ramsey's (1931) procedure for finding events whose subjective probability equals one half. After much experimentation, Davidson et al. selected a six-sided die with the nonsense syllable ZEJ printed on three of its sides and ZOJ on the other three. Subjects didn't seem to care which syllable was associated with the more favorable outcome of a two-outcome bet. Assuming the subjective expected utility model, these "equiprobable" events were used to construct a utility scale, which in turn was used to measure the subjective probabilities of another event. The results of this experiment showed that utility functions and subjective probabilities could be produced for most subjects. Moreover, the resulting SEU values predicted subjects' choices better than did the expected value model.

Later studies by Tversky (1967a) and others have measured utilities and subjective probabilities simultaneously and used them to predict responses to gambles. Although these studies showed that single-attribute utility functions for money could be measured and used to predict people's preferences, this by no means validates expected utility theory. It is, in general, difficult to establish that subjects are using a particular model simply on the basis of that model's ability to predict the outcome of their decision-making process (Birnbaum, 1973; Fischhoff, Goitein & Shapira, 1982; Hoffman, 1960). In this case, moderately good predictability would be guaranteed by the mere fact that people prefer more money to less or higher probabilities of gain to lesser probabilities (Dawes & Corrigan, 1974). Any model capturing this aspect of preference would predict well even if it was seriously deficient on other grounds. Furthermore, subsequent research has shown that the various methods used to determine utility functions may induce substantial biases in those functions, a topic we shall return to later in this chapter.

<u>Tests of the axioms</u>. Many studies have tested utility theory by examining the validity of its axioms. One of the key principles in Savage's axiomatization of subjective expected utility is the <u>extended sure-thing principle</u> which, in one form or another, is crucial to all expected-utility theories. This axiom asserts that if two alternatives have a common outcome under a particular state of nature, then the ordering of the alternatives shall be independent of the value of that common outcome. According to Savage

(1954): "...except, possibly, for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance" (p. 21).

Despite its intuitive appeal, several robust violations of the extended sure-thing principle have been demonstrated. One is the paradox put forth by Allais (1953), which contrasts two hypothetical decision situations, each involving a pair of gambles (expressed in units of one million dollars):

Situation 1. Choose between	Situation 2. Choose between
Gamble 1: 1/2 with probability 1;	Gamble 3: 1/2 with probability .11, 0 with probability .89;
Gamble 2: 2-1/2 with probability .10, 1/2 with probability .89, 0 with probability .01.	Gamble 4: 2-1/2 with probability .10, 0 with probability .90.

Most people prefer gamble 1 to gamble 2, presumably because the small probability of missing the chance of a lifetime to become rich seems very unattractive. At the same time, most people also prefer gamble 4 to gamble 3, presumably because the large difference between the payoffs dominates the small difference between the chances of winning. However, this seemingly innocent pair of preferences is incompatible with utility theory. The first preference implies that

$$.11u(1/2) > .10u(2\text{-}1/2) + .01u(0)$$

whereas the second preference implies the opposite:

$$.10u(2\text{-}1/2) + .01u(0) > .11u(1/2).$$

Another well-known violation of the extended sure-thing principle occurs in the following problem, created by Ellsberg (1961):

Imagine an urn known to contain 90 balls. Thirty of the balls are red; the remaining 60 are black and yellow in unknown proportion. One ball is to be drawn at random from the urn. Consider the following actions and payoffs:

Situation X	30 Red	60 Black	Yellow
Act 1: Bet on red	$100	$ 0	$ 0
Act 2: Bet on black	$ 0	$100	$ 0

If you bet on red, Act 1, you will win $100 if a red ball is drawn and nothing if a black or yellow ball is drawn.

If you bet on black, Act 2, you will win $100 if a black ball is drawn and nothing if a red or yellow ball is drawn.

Now consider the following two actions, under the same circumstances:

Situation Y	30 Red	60 Black	Yellow
Act 3: Bet on red or yellow	$100	$ 0	$100
Act 4: Bet on black or yellow	$ 0	$100	$100

In this problem, the extended sure-thing principle implies that one must choose either 1 and 3 in the two situations, or 2 and 4. Most people select Acts 1 and 4, thus violating the principle. Presumably, they prefer to bet on payoffs whose probabilities are known precisely rather than on payoffs with ambiguous probabilities.

When confronted with such inconsistencies between their intuitions and expected utility theory, some people reject the theory. Or, to use Samuelson's (1950) phrase, they prefer to "satisfy their preferences and let the axioms satisfy themselves." Others reexamine their preferences in the light of the axioms and revise their initial choices.

Savage (1954) offered an illuminating introspective discussion of Allais's example. He admitted that he intuitively preferred gamble 1 to gamble 2 and gamble 4 to gamble 3. Then he adopted another way of looking at the problem: the gambles can be operationalized by a lottery with 100 numbered tickets, one of which is drawn at random to determine the outcome according to the payoff matrix presented in Figure 6. If one of the tickets numbered 12-100 is drawn, it does not matter, in either situation, which gamble is chosen. Hence, one should consider only the possibility that one of the tickets numbered 1-11 is drawn, in which case the two choice situations are identical. Limiting our attention to tickets 1-11, the problem in both situations is whether a 10:1 chance to win 1-1/2 million is preferred to 1/2 a million with certainty. If one prefers gamble 1 to gamble 2, therefore, one should also prefer gamble 3 to gamble 4, if one wishes to be consistent. In concluding his discussion, Savage (1954) wrote:

> It seems to me that in reversing my preference between gamble 3 and 4 I have corrected an error. There is, of course, an important sense in which preferences, being entirely subjective, cannot be in error; but in a different, more subtle sense they can be. Let me illustrate by a simple example containing no reference to uncertainty. A man buying a car for $8,138 is tempted to order it with a radio installed, which will bring the total price to $8,476,

feeling that the difference is trifling. But, when he reflects that, if he already had the car, he certainly would not spend $338 for a radio for it, he realizes that he has made an error (p. 103) [prices revised to reflect inflation since 1954].

Insert Figure 5 about here

Here, Savage used utility theory as a prescriptive framework for ordering his preferences, rather than as a descriptive model of his nonreflective choices. In this spirit, MacCrimmon (1968) presented Allais-type problems to upper-middle-level executives. Although they initially showed the usual inconsistencies, most eventually came to regard their deviations from utility theory as mistakes and desired to correct them. However, Slovic and Tversky (1974) challenged MacCrimmon's procedure for discussing violations on the grounds that it may have pressured the subjects to accept the axioms. Slovic and Tversky presented subjects with arguments for and against the extended sure-thing principle and found persistent violations, even after the axiom was presented in a clear and presumably compelling fashion. Moskowitz (1974) used a variety of problem representations (matrix formats, trees, and verbal presentations) to clarify the principle and still found that it was consistently rejected. MacCrimmon and Larsen (1976) later reevaluated the evidence and suggested that revision of the theory might indeed be in order.

Transitivity of preferences is another key principle of expected utility theory, due to its strong normative appeal and its status as a prerequisite for the existence of any order-preserving utility scale. Transitivity has both a deterministic and a stochastic (probabilistic) form. The latter takes three levels. Let $p(a,b)$ denote the probability that a is preferred to b. If $p(a,b) > 1/2$ and $p(b,c) > 1/2$ then:

strong stochastic transitivity => $p(a,c) > \max[p(a,b),p(b,c)]$;

moderate stochastic transitivity => $p(a,c) > \min[p(a,b),p(b,c)]$;

and weak stochastic transitivity => $p(a,c) > 1/2$.

Tversky (1969) demonstrated violations of weak stochastic transitivity in a situation in which gambles varied in probability and payoff as shown in Table 3. The gambles were constructed so that adjacent gambles had small differences in probability, which subjects tended to ignore when making choices. However, for comparisons between gambles lying far apart in the chain, the cumulative difference in probability of winning (or expected value) dominated the decision. Thus subjects preferred a to b, b to c, c to d, d to e, but e to a, thereby violating transitivity. Tversky's subjects did not realize their preferences were intransitive and some even denied this possibility emphatically.

Insert Table 3 about here

According to Axiom 5 of expected utility theory, any probability mixture of two gambles will always lie between them in the preference order--that is, a mixture can never be better or worse than both the components. Coombs and Huang (1976) tested this "betweenness" property by asking subjects to rank three gambles, a, b, and c in order of attractiveness. Gamble b was a probability mixture of a and c. For example, if a and c represent gambles offering a 50-50 chance to win or lose $1 (for a) or $5 (for c), b would be a four-outcome gamble (-$5, -$1, $1, $5), each outcome having probability .25. Studying decisions among 20 different triples of gambles, Coombs and Huang observed that in a substantial number (about 27%) of the response patterns, the mixture, b, was the most preferred gamble. Coombs and Huang attributed such patterns to subjects' preference for the level of risk embodied in the mixture over the level of risk in the component gambles.

In sum, many of the axioms of utility theory are systematically and consciously violated. These violations have led to a great deal of theoretical activity, both normative and descriptive. On the normative side, a number of decision theorists have proposed revised sets of axioms designed to be consistent with observed behavior without giving up too much of the mathematical convenience and normative value of the earlier models. Thus Chew and MacCrimmon (1979) weakened the substitution axiom in a way that enabled the Allais paradox to be accommodated within the normative theory. Munera and de Neufville (1982) completely did away with the substitution principle and Fishburn (1982b) proposed doing without the transitivity axiom. A review of these and other attempts to revise the von Neumann and Morgenstern axioms is provided by Fishburn (1983).

On the empirical side, the inadequacies of utility theory have stimulated the development of alternative descriptive theories, to which we now turn.

<u>Other Descriptive Theories</u>

<u>Moment theories</u>. Any gamble can be viewed as a probability density distribution over the possible outcomes. The distribution's mean is the gamble's expected value. An early alternative to SEU as a descriptive theory proposed that people base their decisions on the shape of the gamble's distribution as characterized by its first three moments: expected value, variance, and skewness. Fisher (1906) first brought up the potential importance of variance and Allais (1953) used it to criticize expected utility

theory, which assumes that variance preferences can be subsumed under the utility function for money.

For the gamble (x,p,y),

$$\text{variance} = p(1-p)(x-y)^2$$

and

$$\text{skewness} = \frac{1-2p}{\sqrt{p(1-p)}}$$

When probabilities are held constant in two-outcome bets, variance is synonymous with the range of outcomes. Preference for an intermediate level of variance is difficult to account for within utility theory; it suggests a utility curve with several inflection points (alternating regions of risk aversion and risk proneness). Because skewness is monotone with probability in two-outcome bets, preference for a specific skewness level suggests preference for betting at specific probabilities to win and lose.

Edwards (1953; 1954a, b) was the first to study probability and variance preferences. Using two-outcome gambles of equal expected value, he found that 50/50 bets were generally most preferred and bets with .75 probability of winning were avoided. However, his experiment confounded probability differences with variance differences. To remove this confounding, Coombs and Pruitt (1960) constructed a set of two-outcome gambles, all of zero expected value, varying skewness and variance independently. Their subjects exhibited stable probability preferences (usually favoring the highest or lowest probabilities) and variance preferences that interacted with these probability preferences. Specifically, subjects preferred greater variance for gambles containing their preferred probabilities. Although these findings could be explained by an SEU model, Coombs and Pruitt argued that a moment model provided a more parsimonious explanation.

For two-outcome gambles, variance preferences are necessarily confounded with utility and skewness preferences with probability. To break this confounding, Lichtenstein (1965) constructed three-outcome bets that permitted independent variation of probabilities, expected value, variance, and skewness. To assess preferences, she used a bidding method in which subjects stated the largest amount they would pay to play an attractive bet or the largest amount they would pay the experimenter in order to avoid playing an unattractive bet. The results indicated a strong influence of expected value on the amount bid, a slight preference for low variance, and no influence of skewness or probability. The variance effect could be accounted for by a simple utility function.

Another aspect of Lichtenstein's results cast doubt upon the validity of moment theories. This was the finding of large variations in bids within a set of gambles whose first three moment functions were all equal. The case against moment theories was strengthened in subsequent studies by Slovic and Lichtenstein (1968a) and Payne and Braunstein (1971). Slovic and Lichtenstein used specially constructed gambles to manipulate variance without changing the probabilities and payoffs that were explicitly displayed to the subject. The upper half of Figure 6 shows two such bets: a duplex bet and a standard bet, which were termed <u>parallel</u> because both have the same stated probabilities and the same payoffs, namely, .6 chance to win \$2 and .4 chance to lose \$2. Imagine that the bets can be played by spinning pointers on the circular discs shown in Figure 7 such that one wins or loses the amount indicated by the final position of the pointer. To play a duplex bet, one must spin the pointer on both discs. Thus, one can win and not lose, lose and not win, both win and lose, or neither win nor lose. As a consequence, the duplex bet has much less variance than its parallel standard bet. That is, the standard bet leads either to a gain or loss of $2; however, by playing the duplex bet, one has a fairly high probability of breaking even. Most subjects perceived duplex bets and their parallel standard bets as equally attractive, indicating that they were responding to the explicitly stated probabilities and payoffs of the duplex bet and not to its underlying distribution.

Payne & Braunstein used pairs of duplex gambles with equal underlying distributions but different explicit probability values, as illustrated in the lower half of Figure 7. Subjects showed strong preferences for one member of such pairs over the other, which further demonstrates the dominance of explicit or surface information. Taken together, these two studies imply that what passed for probability and variance preferences in earlier studies were byproducts of decision rules applied to the stated probabilities and payoffs. These results helped prompt the study of information processing in decision making, which we will describe later.

Insert Figure 7 about here

Portfolio theory. The failures of the SEU model led Coombs (1969, 1975) to develop an alternative descriptive model called portfolio theory, which highlights risk preference as a determinant of choices among gambles. One assumption of the theory is that preferences among risky alternatives are a function of only two variables, EV and perceived risk (the concept of risk is left undefined, to be inferred from the choice behavior). A second assumption is that the preference function over a set of gambles equal in EV is a single-peaked function of risk.[2] Tests of the theory require additional *ad hoc* assumptions about the subjective definition of risk. For example, Coombs and Meyer (1969) assumed that the perceived risk of a coin-tossing game increases with the denomination of the coin and with the number of tosses one is committed to playing; that is, tossing a coin for a dollar seems at least as risky as tossing for a dime, and tossing for a dollar 10 times seems at least as risky as tossing for a dollar once. Their subjects chose among gambles arranged in sets, with each set having either a fixed monetary denomination or a fixed number of tosses. The results supported the hypothesis that risk, as defined above, was a determinant of subjects' preferences.

In another attempt to infer the meaning of risk and validate portfolio theory, Coombs and Huang (1970a) assumed that a probability mixture of two gambles, each with the same expected value, will have a level of risk intermediate between the two. They hypothesized that individuals with single-peaked preference functions over risk would, for some mixtures, violate the betweenness principle of utility theory. The data confirmed their hypothesis.

Portfolio theory has stimulated a number of empirical and theoretical studies of the perceived risk of a gamble. Coombs and Huang (1970b) examined the effects on risk judgments of various transformations of probabilities and payoffs and Coombs and Bowen (1971) examined the effects of variance and skewness. The results suggested that risk could be predicted by a simple polynomial model incorporating variables such as expectation and skewness. Luce (1980, 1981) suggested four specific models of a gamble's risk, depending upon how risk changes given a change of scale and how a probability density distribution aggregates into a single value for risk. Weber (in press) found that multiplying all outcomes by a scale factor had an additive effect on perceived risk, thus supporting one of Luce's models. Coombs and Lehner (1981) argued that models involving moments of distributions and functions of the moments imply a symmetry that does not correspond with people's perceptions of risk. For example, adding \$10 to the positive outcome of a gamble has less effect on perceived risk than a symmetric addition of \$10 to the negative outcome. They proposed separating the good and bad outcomes by means of a model that does not include moments. If g is a gamble with a winning and a losing outcome, W and L, having probabilities of p and q, respectively, and the residual probability (1-p-q) is associated with a zero outcome, the Coombs and Lehner model for the riskiness of g is:

$$R(g) = \theta_1(p)\theta_2(W) + \theta_3(q)\theta_4(L)$$

where the θ_i are real-valued monotone functions.

Prospect theory. Kahneman and Tversky (1979) also developed an algebraic model that was designed to remedy the descriptive failures of utility theory. Referring to gambles as risky prospects, they named their approach prospect theory. They demonstrated three pervasive phenomena, labeled the certainty effect, the reflection effect, and the isolation effect, that were used to determine the structure of prospect theory.

The certainty effect is the tendency to overweight outcomes that are considered certain, relative to outcomes that are merely probable. For example, most subjects preferred option B over option A in the following choice (outcomes were Israeli pounds):

A: Win 4,000 with probability .80

B: Win 3,000 with certainty

However, the majority of subjects also preferred C over D:

C: Win 4,000 with probability .20

D: Win 3,000 with probability .25

This pattern of preferences, B over A and C over D, violates a lemma of utility theory's substitutability axiom. Option C is a probability mixture of A (i.e., C can be seen as a .25 chance to win A) and D is the same probability mixture of B. Thus if B is preferred over A, D must be preferred over C. Kahneman and Tversky explained the violation as a result of the certainty effect; the reduction in probability has greater impact when it alters the character of the prospect from a sure gain (B) to a probable one (D) than when both the original and the reduced prospects (A and C) are uncertain.

The reflection effect is the tendency for preferences between positive prospects to be reversed when the gains are replaced by losses. For example, when options A, B, C and D, above, are changed to losses (e.g., A': lose 4,000 with probability .80), the majority of subjects prefer A' over B' and D' over C'. The reflection effect implies risk aversion in the positive domain and risk proneness in the negative domain.

The isolation effect is the tendency to simplify choices between alternatives by disregarding components that the alternatives share and focusing on the components that differentiate them (Tversky, 1972b). This tendency can produce inconsistent preferences when a pair of prospects are decomposed into common and distinctive components in different ways. For example:

Consider a two-stage game in which stage 1 offers a probability of .75 to end the game without winning anything and a probability of .25 to move to the second stage. The second stage offers a choice between:

E: Win 4,000 with probability .8, and

F: Win 3,000 with certainty.

One's choice for the second stage must be made before the game starts.

In terms of the final outcomes, the real choice in this game is between a .25 x .80 = .20 chance to win 4,000 and a .25 x 1.0 = .25 chance to win 3,000, as with options C and D above. There most subjects chose option C. However, the isolation effect leads subjects to disregard the common first stage and treat the game as if it were composed only of the second stage: 78% of the subjects chose option F.

The preceding two-stage game shows that preferences can be changed by different representations of probabilities. The following pairs of choices show the isolation effect with varying representations of outcomes:

Pair 1: In addition to whatever you own, you have been given 1,000. You are now asked to choose between:

G: win 1,000 with probability .50, and

H: Win 500 with certainty.

Pair 2: In addition to whatever you own, you have been given 2,000. You are now asked to choose between

I: lose 1,000 with probability .50, and

J: lose 500 with certainty.

The majority choices were H and I. However, the final states are the same in the two pairs; G = (2,000, .50, 1,000) = I and H = 1,500 = J. Evidently, subjects did not integrate the bonuses with the respective prospects. These results are inconsistent with utility theory, which assigns the same utility to a final outcome regardless of how that outcome was obtained.

Prospect theory was designed to accommodate these various behavioral phenomena. The theory distinguishes two phases in the choice process. An editing phase organizes and reformulates the options so as to simplify subsequent evaluation and choice. The major operations of editing include: (a) coding of outcomes as gains or losses around some neutral reference point, (b) combination of probabilities associated with identical outcomes, (c) separation of risky from riskless components, and (d) cancellation of components shared by all offered prospects.

Following editing, the decision maker is assumed to evaluate each of the edited prospects and to choose the prospect with the highest value. The overall value of an edited prospect, denoted V, is expressed in terms of a value function $v(x)$, which attaches a subjective worth to each outcome, and a weighting function, $\pi(p)$, which expresses the subjective importance attached to the probability of obtaining a particular outcome. The attractiveness of a gamble that offers a chance of p to gain x and a chance of q to lose y is:

$$V = \pi(p)v(x) + \pi(q)v(y)$$

A slightly different equation is applied if both outcomes are on the same side of the zero point.

The decision weight $\pi(p)$ is a monotonic function of p but is not a probability. Although $\pi(0) = 0$ and $\pi(1) = 1$, the function is not well behaved near the end points. In addition, for low probabilities, $\pi(p) > p$ but $\pi(p) + \pi(1-p) < 1$. Also, $\pi(pq)/\pi(p) < \pi(pqr)/\pi(pr)$ for all $0 < p,q,r < 1$. That is, for any fixed probability ratio, q, the ratio of decision weights is closer to unity when the probabilities are lower than when they are high. For example, $\pi(.4)/\pi(.8) > \pi(.1)/\pi(.2)$. An illustrative weighting function that satisfies these properties is shown in Figure 8.

Insert Figure 8 about here

The general form of the value function is shown in Figure 9. It is defined on gains and losses relative to some psychologically neutral reference point. The function is steeper for losses than for gains, implying that a given change in one's status hurts more as a loss than it pleases as a gain. A third feature is that it is concave above the reference point and convex below it, meaning, for example, that the subjective difference between gaining (or losing) $10 and $20 is greater than the difference between gaining (or losing) $100 and $120.

Insert Figure 9 about here

Payne, Laughhunn, and Crum (1980, 1981) treated the reference point as a level of aspiration. They showed how equal additions or subtractions to all outcomes in a gamble, by changing the relation of the outcomes to the level of aspiration, could markedly affect an individual's preferences.

<u>Framing</u>. The editing phase of prospect theory was subsequently labeled <u>framing</u> by Tversky and Kahneman (1981). Much as changes in vantage point induce alternative perspectives on a visual scene, the same decision problem can be viewed in alternative frames. The frame that is adopted is determined in part by the external formulation of the problem and in part by the standards, habits, and idiosyncratic perspectives of the decision maker.

If π and v were linear functions, preferences among risky prospects would be independent of the problem frame. However, because of the nonlinearity of these functions, normatively inconsequential changes in the problem frame can have profound effects on preferences. This was illustrated in several of the examples described above. Tversky and Kahneman (1981) provided additional examples of framing effects, such as the following pair of problems, given to separate groups of respondents.

Problem 1. Imagine that the U.S. is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the consequences of the programs are as follows: If Program A is adopted, 200 people will be saved. If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. Which of the two programs would you favor?

Problem 2. (Same cover story as Problem 1) If Program C is adopted, 400 people will die. If Program D is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. Which of the two programs would you favor?

Although the two problems are formally identical, the preferences tend to be quite different. In a study of college students, 72% of the respondents chose Program A over Program B, whereas 78% chose Program D over Program C. A framing interpretation of this inconsistency is that the "save lives" wording of the first problem induces a reference point of 600 lives lost, whereas the "people will die" wording of Problem 2 induces a reference point of no lives lost. Thus Problem 1 falls in the concave gain region of the value function, whereas Problem 2 is in the convex loss region. In another study, physicians and patients reversed their choice between radiation vs. surgical treatments for lung cancer when the relevant statistics were changed from probabilities of surviving for various lengths of time after treatment to probabilities of not surviving (McNeil, Pauker, Sox & Tversky, 1982).

One class of framing effects, called "pseudocertainty" by Tversky and Kahneman (1981), converts uncertainty about an outcome to certainty about a subset of that outcome. Protective actions, for example, may be easily manipulated so as to vary the apparent certainty with which they prevent harm. Thus, an insurance policy that covers fire but not flood could be presented either as a reduction in the overall probability of property loss or as full protection against the specific risk of fire. Because possible outcomes are undervalued in comparison with certain outcomes, Tversky and Kahneman hypothesized that such an insurance policy would appear more attractive in the context that offers unconditional protection against a restricted set of problems.

Slovic, Fischhoff and Lichtenstein (1982) found empirical support for this conjecture in the context of one particular kind of insurance, vaccination. Two forms of a "vaccination questionnaire" were created. Form I (probabilistic protection) described a disease expected to afflict 20% of the population and asked people whether they would volunteer to receive a vaccine that protects half of the people receiving it. In Form II (pseudocertainty), there were two mutually exclusive and equiprobable strains of the disease, each likely to afflict 10% of the population; the vaccination was said to give complete protection against one strain and no protection against the other. More people were willing to be vaccinated with Form II than with Form I.

In order to use prospect theory predictively, one must be able to discern the frames that subjects adopt. In the examples given above, investigators were successful in inducing subjects to adopt frames whose interpretation was clear to other investigators. Using similar kinds of problems, however, Fischhoff (1983) was less successful. Each problem offered a choice between a sure loss (e.g., $50) and a gamble with two negative outcomes (e.g., a 50-50 chance to lose $40 or $60). Three possible frames were identified, two of which would, according to prospect theory, lead to the choice of the gamble, whereas the third would lead to choice of the the sure thing. Most subjects preferred the gamble, an example of risk seeking in the domain of losses that prospect theory would predict. Most subjects also indicated that they found the frames leading to the gamble (assuming the truth of prospect theory) to be "most natural" and that they had, in fact, used those frames. However, there was no relationship between individual subjects' choices and their frame preferences. Nor was it possible to alter the choice patterns by changing the wording of the problems so as to highlight particular frames. These results suggest: (a) that people may not be able to introspect well regarding the judgmental processes involved in framing and (b) that some natural frames may be so robust that it is difficult to dislodge them by experimental manipulation. The study points out the need for a substantive theory showing how framing works in particular situations.

Regret theories. The demonstrated violations of utility theory that prompted prospect theory have also led to the development of other theories of decision making under risk. Bell (1982), Loomes and Sugden (1982), and Sage and White (1983) have developed highly similar theories based on regret. In these theories, the utility of a risk prospect depends not only on the choiceless or inherent utility of each possible outcome but also on the regret or rejoicing one might experience upon making a choice, receiving an outcome, and comparing the actual outcome with the outcome that would have obtained had some other choice been made. Thus a preference for A over B really means "choosing A and rejecting B is preferable to choosing B and rejecting A."

Since every choice depends on the regret/rejoicing involved in the particular pair of options being considered, transitivity of choice is not ensured in regret theories. Indeed, Loomes and Sugden (1982) argued against transitivity as an appropriate prescriptive goal.

Another principle of utility theory that is rejected by regret theories is that options with identical probability distributions of outcomes are equivalent. Consider the following two formally identical gambles. The chosen gamble will be decided by blindly drawing a ball from an urn containing one red, one white, and one black ball.

	Red	White	Black
Option A	Win $8	Win $2	Win $1
Option B	Win $1	Win $8	win $2

The regret/rejoicing function is not linear with differences in outcomes (if it were, regret theories would be identical to utility theory). Thus a person may have a strong preference for one of the two options, since if A is chosen the rejoicing resulting from drawing the red ball and winning $8 rather than $1 would not necessarily balance the sum of the regrets felt from drawing the white ($2-$8) or black ($1-$2) ball.

With suitable assumptions about the shapes of the choiceless utility and regret functions, regret theories can explain many, but not all, of the best-known violations of utility theory.

Expected Utility Theory: An Assessment

Experimental psychology has played a key role in testing the descriptive validity of utility theory. The insights gained from empirical studies have been synthesized in descriptive theories in ways that dramatize the major differences between the normative postulates of the theory and the behaviors that people exhibit when making decisions among risky alternatives.

Examination of trends across thirty years of empirical research shows an increasing sensitivity to psychological considerations. The regret theories certainly fit this pattern, as does prospect theory with its emphasis on problem representations, reference points, and editing operations designed to simplify information processing. The increasing prominance of information-processing considerations is a theme we shall return to later in this chapter. In fact, the impact of information-processing considerations appears so great that even the proponents of prospect theory and regret theory see their models as approximate, incomplete, and much too simplified descriptions of behavior.

These descriptive considerations have also affected the normative theory. From Bernoulli to Savage, the history of utility theory is one of successive adjustments, incorporating increasing subjectivity into the theory, in order to accommodate observed behavior. The new descriptive theories are likely to leave their mark on normative theory, too, although the nature of that mark remains to be seen. More thought is needed regarding whether the effects of regret, decision weights, reference points, and framing should be treated as errors of judgment or as valid elements of human experience whose appeal remains after thoughtful deliberation. In time, some of the now-sacred axioms may lose some of their normative status because of their descriptive shortcomings. What is evident, however, is that the field of risky decision making, so simple, orderly, and "wrapped up" two decades ago, has been shaken into an exciting state of turmoil by recent experimental results.

Although utility theory is obviously an incorrect descriptive model, it continues to play a central role in many important practical problems ranging from counseling individuals on family planning decisions to guiding government policies (Feather, 1982; Fischhoff, Lichtenstein, Slovic, Derby & Keeney, 1981). Coombs (1980) observed that "the vitality of [utility] theory in the face of criticism is astounding" (p. 346-347). One reason for the theory's staying power is that it provides an excellent approximation to many judgments and decisions. Goodman, Saltzman, Edwards, and Krantz (1979) found that even the simple expected value model accounted for 88% of the variance of maximum buying prices for gambles in a study conducted in a Las Vegas casino. Coombs and Huang (1976), using gambles specially constructed to induce violations of the betweenness principle, found that utility theory still accounted for 86% of the preference orderings. As Coombs (1980) concluded, "A theory that provides good approximations, even though it is wrong in principle, is going to be used until a more useful theory comes along" (p. 348).

The problem, of course, is that effective personal and social decisions often require more than approximate understanding of behavior. For example, Kunreuther et al. (1978) showed that flood insurance programs, designed under the assumption that people maximize expected utility when buying insurance, failed because those decisions were actually made on other grounds.

Single-Attribute Probabilistic Choice Models[3]

Consider a decision situation in which there is no uncertainty about the outcomes; you get what you choose. An intriguing characteristic of choices in such situations is their inconsistency. People do not always make the same choice when faced with the same alternatives under seemingly identical conditions. Inconsistency remains even after controlling for factors such as learning, satiation, or changes in taste over time.

One way of accounting for inconsistency is by postulating a random element in choice, replacing the deterministic notion of preference with a probabilistic one. In modeling such choices, the (absolute) preference of x over y is replaced by the probability of choosing x over y, denoted p(x,y). This probability can be estimated from the relative frequency with which x is chosen over y. It is commonly viewed as a measure of the degree to which x is preferred over y. If an individual strongly prefers x to y, one would expect p(x,y) to be close to unity and hence p(y,x) to be close to zero. If the individual is indifferent between x and y, one would expect p(x,y) to be close to one half. The deterministic notion of preference is viewed, in this framework, as a special case where all pairwise choice probabilities are either zero, one, or one half.

Probabilistic decision theories can be divided into two types: constant utility models and random utility models. Constant utility models assume that each alternative has a fixed utility value and that the probability of choosing one alternative over another is a function of the distance between their utilities. The decision task is viewed as a discrimination problem in which the individual is trying to determine which alternative would be more satisfying. The greater the distance between the utilities, the easier the discrimination. Constant utility models resemble psychophysical theories in which, for example, the probability of judging one object as heavier than another is expressed as a monotonic function of the difference between their weights.

Random utility models assume that decision makers always choose the alternative that has the highest utility, but the utilities themselves are random variables rather than constants. The actual choice mechanism, therefore, is purely deterministic, but the utility of each alternative varies from moment to moment.

Thurstone's Random Utility Model

One of the first to recognize the probabilistic nature of choice behavior was Thurstone (1927):

> An observer is not consistent in his comparative judgments from one occasion to the next. He gives different comparative judgments on successive occasions about the same pair of stimuli. Hence we conclude that the discriminal process corresponding to a given stimulus is not fixed. It fluctuates (p. 271).

To explain these fluctuations, Thurstone introduced the law of comparative judgment. This model represents stimuli as distributions, or random variables, reflecting the momentary fluctuations of their perceived values. The probability of choosing one alternative over another is the probability that the first random variable exceeds the second. Thurstone proposed several alternative sets of assumptions about the form and the interrelations among the stimulus distributions. In the simplest case, called Case V, all the distributions are independent and normal with equal variance. In this case, the means of the distributions can be easily calculated from the observed choice probabilities. Thurstone used his model to scale preferences among foods, potential birthday gifts, and other riskless options.

Coombs' Random Utility Model

Coombs (1958) proposed that both the choice alternatives and the decision maker's ideal point could be represented as random variables along a common underlying dimension. As in the Thurstonian model, the distributions are independent of each other and unimodal. Because both the alternatives and the ideal point are random variables, so is the distance between them. The probability of preferring one alternative over another, therefore, equals the probability that the distance between one alternative and the ideal is less than that between the other alternative and the ideal.

Preference for room temperature may illustrate the model. One has an ideal temperature level that fluctuates over time. One's perception of any given temperature also fluctuates randomly. In comparing two temperature conditions, one choses the temperature that appears closer to the ideal point at the moment of comparison.

Coombs tested his model via its implications for transitivity. The model implies that the kind of transitivity that will be observed depends on the position of the alternatives relative to the ideal point. Strong stochastic transitivity is expected to be satisfied when all three alternatives are on the same side of the ideal. When the ideal point is between the stimulus points, its variability should combine with variability of the stimuli to produce inconsistency and a greater amount of intransitivity. These predictions about transitivity were supported in a study using as stimuli 12 shades of gray varying only in brightness (Coombs, 1958). The subjects, given all possible sets of 4 stimuli, were asked to choose among them according to how well each represented the notion of an "ideal" gray. Further support for the model's transitivity predictions was described by Coombs (1964).

Coombs' work is significant because it demonstrates that choice probabilities cannot be converted into differences along a single dimension without considering the relationships between the alternatives and the ideal point. Thus, it points to the importance of option comparability in determining preference.

Luce's Constant Utility Model.

Instead of making assumptions about the form of the value distribution (which are typically hard to justify), Luce (1959) assumed that choice probabilities satisfy one simple but powerful axiom. Despite the different conceptualization of choice embodied in Luce's model and Case V of Thurstone, the two models have been shown to be closely related (see, e.g., Block & Marschak, 1960, and the chapter by Luce and Krumhansl in this volume).

Suppose that all choice probabilities are neither zero nor one. Let T be a finite set of alternatives and R be any subset of T. Luce's choice axiom asserts that the probability of choosing an element x of R, from the entire set T, p(x;T), equals the probability that the selected alternative will be in the subset R, p(R;T), multiplied by the probability of choosing x from R, p(x;R). That is,

$$p(x;T) = p(x;R)p(R;T) \quad \text{for} \quad R \subset T.$$

Thus, the probability of selecting roast beef (x) from an entire menu (T) equals the probability of selecting roast beef from the meat entrees (R) times the probability of choosing a meat entree.

Let p(x,y) be the probability of choosing x over y. If p(x,y) = 0 for some x, y in T, it is further assumed that x can be deleted from any choice set containing y without affecting the choice probabilities. Doing this allows one to reduce all choice problems to the imperfect discrimination case where all probabilities differ from zero or one.

Using the choice axiom and the usual laws of probability theory, Luce derived a number of testable consequences. One is the constant ratio rule:

$$\frac{p(x;T)}{p(y;T)} = \frac{p(x;R)}{p(y;R)} = \frac{p(x,y)}{p(y,x)}$$

That is, ratios of the form $p(x;R)/p(y;R)$ are independent of R. Thus the odds of choosing steak rather than roast beef for dinner are the same for all menus containing both entrees.

The probability of choosing any item x from T can be found from the pairwise choice probabilities via the relationship:

$$p(x;T) = \frac{1}{\sum_{y}[p(y,x)/p(x,y)]}$$

Luce's model implies the existence of a ratio scale of preference. To construct such a scale, select one element, a, and set its value at one,

$$v(a) = 1.$$

Then for any other member, x, of the choice set:

$$v(x) = \frac{p(x,a)}{p(a,x)} = \frac{p(x;T)}{p(a;T)} .$$

Although the choice axiom is directly testable, many observations are needed in order to obtain stable estimates of choice probabilities. Consequently, most studies combine the choices of several individuals in estimating the probabilities. Unfortunately, the model cannot be unambiguously tested with group data, because each individual in the group may satisfy the axiom, yet the average probabilities violate it, or vice versa.

Within the constraints of these methodological difficulties, the choice axiom has been tested in a wide variety of settings ranging from consumer and political choices to studies of learning and psychophysics, with animals as subjects as well as humans. After a comprehensive review of these studies, Luce (1977) reached a conclusion similar to that given above for expected utility theory. The choice axiom is surely incorrect in many settings, yet it often provides a reasonable approximation. Moreover, the choice axiom embodies, in probabilistic form, an important principle of rational choice put forth by Arrow (1951): decisions should be independent of irrelevant alternatives. Hence, the axiom often serves as a basis for rational, probabilistic theories of economic and social behavior.

The inadequacies of the choice axiom are most clearly seen in violations of simple scalability (Luce & Suppes, 1965). This property, which holds for all constant utility models, requires the alternatives to be scaled so that each choice probability is expressable as a monotone function of the scale values of the respective alternatives. Coombs' (1958) data on preferences among shades of gray (discussed above) violate simple scalability. Another violation of this principle comes from Debreu (1960a):

Suppose you are offered a choice among three records: a suite by Debussy, denoted D, and two different recordings of the same Beethoven symphony, denoted B_1 and B_2. Assume that the two Beethoven recordings are of equal quality and that you are indifferent between adding a Debussy or a Beethoven to your record collection. Hence,

$$p(B_1;B_1,B_2) = p(D;D,B_1) = p(D;D,B_2) = 1/2.$$

It follows from Luce's model that

$$p(D;D,B_1,B_2) = 1/3.$$

Intuitively, however, the basic conflict between Debussy and Beethoven is not likely to be affected by adding another Beethoven recording. Instead, Debreu suggested that B_1 and B_2 be treated as one alternative to be compared with D. Consequently, one would expect $p(D;D,B_1,B_2)$ to be close to one-half, whereas $p(B_1;D,B_1,B_2) = p(B_2;D,B_1,B_2)$ will be close to one-fourth, contrary to simple scalability. Empirical support for Debreu's hypothesis was presented by Becker, DeGroot, and Marschak (1963) in a study of choice among gambles. Although offered as a criticism of Luce's model, Debreu's example applies to any model based on simple scalability.

A minor modification of an example due to L. J. Savage (see Tversky, 1972b) illustrates yet another difficulty encountered by simple scalability. Imagine a person who is indifferent between a trip to Paris and a trip to Rome, so that P(Paris, Rome) = 1/2. When the person is offered a new alternative consisting of the trip to Paris plus a $1 bonus, denoted Paris +, this option will certainly be preferred it to the original trip to Paris, so p(Paris +, Paris) = 1. Simple scalability predicts that p(Paris +, Rome) = 1, which is counterintuitive. It is unlikely that a relatively small bonus would resolve the conflict so completely. Rather, p(Paris +; Rome) should remain close to 1/2. Experimental data (e.g., Tversky & Russo, 1969) support this intuition. Choice probabilities, therefore, reflect not only the utilities of the alternatives in question, but also the difficulty of comparing them. Thus, an extreme choice probability (i.e., close to 0 or 1) can result from either a large discrepancy in value or from an easy comparison, as in the case of the added bonus. The comparability of the alternatives, however, cannot be captured by their scale values; hence simple scalability is violated.

Multi-Attribute Probabilistic Choice Models

We turn now to decision-making models that consider the more complex situation in which the objects or actions have several aspects, attributes, or dimensions. For the sake of continuity, we begin with probabilistic choice models, thus continuing directly from the immediately preceding text.

Choice and Similarity

As the preceding example suggests, simple scalability seems particularly unlikely to hold when the choice set contains similar outcomes. As a consequence, choice models may be restricted to sets of dissimilar alternatives or elaborated to incorporate the structure of the choice set. Two theorists, Restle (1961) and Tversky (1972a, b; Tversky & Sattath, 1979) have followed this latter course.

Restle proposed a variation of Luce's theory that assumed choices depend on the elements that differentiate the alternatives, rather than on those common to them. Each alternative is viewed as a set of objects or outcomes, as schematized in Figure 10. The set actually contributing to the choice of A is the set difference, $A \cap \bar{B}$, denoted A-B. The set contributing to the choice of B is B-A. The intersection of A and B, marked I, is assumed not to affect the response probabilities. In Restle's theory,

$$P(A,B) = \frac{V(A-B)}{V[(A-B) \cup (B-A)]}$$

where V is a value measure across all the elements of the set. If the sets A and B have no overlap, then Restle's and Luce's theories make equivalent predictions about choice probabilities.

Insert Figure 10 about here

Rumelhart and Greeno (1971) compared the Luce and Restle models. Their stimuli were nine famous people of the era, including three political leaders, three athletes, and three film stars. Subjects were presented with pairs of these people and were asked to choose the one with whom they would prefer to spend an hour. It was assumed that there were important overlaps or similarities within but not across categories. The Restle model, with parameters that took similarity into account, predicted the choices between similar pairs much better than did the Luce model.

Elimination by Aspects

Tversky (1972a, b) developed a context-dependent probabilistic model that generalized both the Luce and Restle theories. Tversky's model, called elimination by aspects (EBA), views each alternative as a collection of measurable (i.e., scalable) aspects and describes choice as a covert process of successive elimination. At each stage in the process, one selects an aspect of the available alternatives with probability proportional to its value. Any alternative that does not include the chosen aspect is eliminated and the process continues until a single alternative remains. Consider, for example, the choice of a restaurant for dinner. The first aspect selected may be seafood, thus eliminating all restaurants that do not serve seafood. Another aspect, say a price range, is then selected and alternatives failing to met this criterion are eliminated. The process continues until only one restaurant, which includes all the selected aspects, remains.

Formally, the EBA model is represented by a recursive formula, which expresses the probability of choosing option x from set A as a weighted sum of the probabilities of choosing x from proper subsets of A. For binary choices, EBA coincides with Restle's model (and hence with Luce's if the aspects for each alternative are disjoint). EBA predicts violations of simple scalability for overlapping alternatives and makes sensible predictions for the choice problems posed by Debreu and Savage. Moreover, it has several testable consequences that considerably constrain the observed choice probabilities and permit a measurement-free test of the model.

The EBA model is an appealing way to make decisions because it is easy to apply and is easy to explain and justify to others. There is no guarantee, however, that the model will lead to normatively defensible decisions. That is, the EBA process cannot ensure that the alternatives retained are superior to those that are eliminated.

<u>Preference trees</u>

The EBA model imposes no restrictions on the structure of the choice aspects. As a result, for a choice set of n alternatives, a large number of scale values (2^n-2, corresponding to the number of proper subsets of the total set) are required to fit the model. Thus, the model cannot be estimated from binary choice probabilities because the number of parameters exceeds the number of data points. Tversky and Sattath (1979) simplified the EBA model by imposing some structure on the set of aspects in a way that reduced the number of parameters to 2n-2. They represented choice alternatives by a tree-like graph in which each terminal node is associated with a single alternative and each link between nodes is associated with the set of aspects that are shared by all the alternatives that include or follow from that link and are not shared by any of the alternatives that do not include that link. The length of each link in the tree represents the value measure of the respective set of aspects. Hence, the set of all aspects that belong to a given alternative is represented by the path from the root of the tree to the terminal node represented by the alternative and the length of the path represents the overall measure of the alternative.

Insert Figure 11 about here

Figure 11 illustrates a tree representation of a menu. The set of alternatives consists of five entrees, which appear as the terminal nodes of the tree. Thus, the link labelled λ represents the aspects shared by all meat entrees but not by fish, θ represents the aspects shared by steak and roast beef but not by lamb or fish, and γ represents the unique aspects of lamb. The suggested labels of the clusters (defined by the links) are displayed horizontally. When choices are represented in this way, elimination by aspects reduces to elimination by tree (EBT). That is, one selects a link from the tree (with probability proportional to its length) and then eliminates all alternatives that do not include the selected link. This process is then applied to the selected branch until only one alternative remains. Tree models have provided excellent representations for a wide variety of data, including choices among the nine celebrities studied by Rumelhart and Greeno (1971) and preferences among political parties and among academic disciplines collected by Sjöberg (1977).

Probabilistic Theories of Choice: An Assessment

The development of probabilistic theories of choice shows some remarkable parallels with the development of risky decision models. Both lines of research show a progression from simple conceptions to more complex psychological models. Early models, such as SEU and the constant ratio rule, ignored framing or contextual manipulations; later models, such as prospect theory and preference trees, did not. The introduction of psychological complications, such as reference points and regret, into theories of decision making under risk are paralleled by the introduction of comparability and similarity parameters in theories of riskless choice. Nonetheless, even the most elaborate descriptive theories in each domain are viewed by their creators as useful approximations, but incomplete and inadequate.

Multi-Attribute Value Models

There are several models for preference among multi-attribute objects that entirely disregard both uncertainty about the state of the world that will obtain and uncertainty about one's preferences. Following Keeney and Raiffa (1976), we call these riskless algebraic utility models <u>value</u> models, saving the term <u>utility</u> for the risky models.[4] In all these models, all objects are assumed to be described by a common set of dimensions or attributes.

Axiomatized models

The axioms for these models, like the axioms for Expected Utility Theory, specify the conditions under which a value function exists and point to the measurement procedures by which the value functions may be found.

<u>Weak order</u>. The simplest model, with the fewest assumptions, leads to a value function that is only ordinally scaled. The key axiom for this <u>weak</u>

order model (Krantz, Luce, Suppes & Tversky, 1971) is transitivity. The model states that

$$A \succsim B \text{ iff } v(A) \geq v(B),$$

where v is an ordinal function. One measurement procedure for implementing this model is ranking, with the ranks constituting the value function. A second method is indifference curve construction, which places the objects in an n-dimensional indifference map (where n is the number of attributes that describe each object). Figure 12 shows a prototypical indifference map for two unspecified attributes, 1 and 2. Each object is represented by a point in the space. Indifference curves connect points that are equal in value. Every point above and to the right of a given indifference curve is preferred to any point on the curve, whereas any point to the left and below the indifference curve has less value. Although indifference maps can, in theory, be derived for objects with more than two attributes, the judgments become so difficult that the procedure is rarely undertaken. Consider, for example, filling in the blank in dinner B, shown in Table 4, such that you are indifferent between dinner A and dinner B.

Insert Table 4 and Figure 12 about here

Stronger models. In order to facilitate measurement, additional assumptions are made, creating stronger models. Some additional axioms ensure interval scaling for the overall value; others, called decomposition models, yield interval scales of value for each of the attributes and a rule for combining the separate values into an overall value for each object (von Winterfeldt, 1975).

One of the strongest value models is the additive conjoint model:

$$A \succsim B \text{ iff}$$

$$v(A) \geqslant v(B), \text{ where}$$

$$v(A) = \sum_{i=1}^{n} v_i(A_i).$$

That is, under the axioms (not given here; see Debreu, 1960b and Keeney & Raiffa, 1976), the desirability of an object is determined by the sum of the values of its attributes. One of the strong assumptions made by this model is that the attributes are <u>pairwise preferentially independent</u>, which means that the tradeoff between any two attributes does not depend on the levels of any other attribute.

In order to find the value functions, v_i, the decision maker must make several direct tradeoffs between every pair of attributes. These tradeoff questions are called <u>dual standard sequences</u>. For example:

> Holding constant all other attributes, suppose an apartment has a rent of $420 and distance to work of 5 miles. How many more miles would you be willing to drive to work if the rent decreased to $400?

The test for pairwise preferential independence requires several more such tradeoffs as the levels of all the other attributes are systematically varied.

<u>Non-axiomatized value models</u>

Axiomatized value models have no error component. People are assumed to have stable preferences they can reliably report. However, for complex problems, common sense (as well as much of the research reviewed elsewhere in this chapter) suggests that the axiomatically justified measurement techniques may generate enough cognitive strain to cause serious errors. Thus several authors (e.g., von Winterfeldt, 1975) have argued that more accurate representations of people's preferences may be obtained with very simple assessment procedures even if these procedures lack axiomatic justification.

SMART. One such simplified multi-attribute value model, proposed by Edwards (1977), is called SMART (Simple Multiattribute Rating Technique). Edwards defended SMART by arguing:

> ...While it lacks...theoretical elegance..., it has the great advantage of being easily taught to and used by a busy decision maker.... Moreover, it requires no judgments of preference or indifference among hypothetical entities. My experience with elicitation procedures suggests that such hypothetical judgments are unreliable and unrepresentative of real preferences; worse, they bore untutored decision makers into either rejection of the whole process or acceptance of answers suggested by the sequence of questions rather than answers that reflect their real values. (p. 327)

SMART reduces the elicitation procedure to 10 relatively easy steps, the most significant of which are:

- Rate the attributes by assigning the number 10 to the least important attribute and assigning numbers to the others that preserve the ratio of importance to the least important (and, as a check, ratios of all pairs of attributes).

- Normalize the importance weights by dividing each by the sum of the weights. Call these normalized weights w_j.

- Rate the value of each object, i, on each attribute, j, using a scale from 0 to 100, with 0 defined as the worst plausible value and 100 defined as the best plausible value. Call these ratings v_{ij}.

- Calculate $V_i = \sum_j w_j v_{ij}$ for each object.

• Choose the object with the largest v_i.

With this model, the resultant values are scaled in a common range and the weights reflect the relative importance of the attribute dimensions. In contrast, axiomatically-based value models either (a) do not elicit weights directly because the tradeoffs across attribute dimensions emerge from the value scaling procedures (as with dual standard sequences for the additive conjoint model) or (b) have indirectly assessed weights that are uninterpreted (as with the multiplicative model not described here; see Keeney & Raiffa, 1976).

<u>Social judgment theory</u>. Social judgment theory is a set of techniques, based on the previously described lens model, designed to aid societal decision making (see Hammond & Adelman, 1976; Hammond et al., 1975; and Hammond, Rohrbaugh, Mumpower & Adelman, 1977). Unlike decision theories, social judgment theory does not prescribe a course of action. Instead, it is used to externalize decision makers' implicit policies (i.e., their preferences and beliefs). In applications of the theory, the judgments of another person or another social group replace the criterion on the other side of the lens. Thus the technique is useful for locating sources of conflict. The art of the social judgment theorist lies in designing stimulus profiles that appropriately represent the configurations of possible alternatives. An armamentarium of on-line computer programs enables decision makers to respond to stimulus profiles designed by the analysts and to receive feedback about the weights and the functional relationships implied by their judgments.

The application of social judgment theory is illustrated by a study designed to help the Denver Police Department select the best bullet to use (Hammond, Stewart, Adelman & Wascoe, 1975). The analysts asked ballistics experts to rate available bullets on the dimensions deemed important by the decision makers: stopping power, injury potential, and danger to bystanaders. The analysts then asked the decision makers for wholistic judgments of preference for bullet profiles drawn from the experts' ratings. Correlational analyses based on the lens model were used to infer the importance weights for each of the attribute dimensions. The analysis revealed a bullet that had greater stopping power (which the police wanted) yet no greater potential for injury and less danger to bystanders than the bullet the police were then using. This bullet was subsequently adopted for use. Social judgment theory has also been used to facilitate public input into regional planning, to structure faculty participation in university policy decisions, and to reduce conflict in labor-management negotiations (Hammond et al., 1977).

Validating value models

In a sense, the results from axiomatized value models are automatically valid. If you accept the axioms and are satisfied with your judgments, then the resulting evaluations follow logically, as reflections of your values. Furthermore, the consistency checks (e.g., transitivity) built into some of the elicitation procedures allow you to identify and correct errors. But you may not be fully satisfied with your judgments. How good then is the model based on them? In the realm of values, there is no public "right answer" against which to compare a model's output. Non-axiomatic models face the same problem, without the partial reassurance of asking normatively justified questions.

The validity of the lens model that underlies social judgment theory has been tested in settings in which some criterion exists. Linear additive models derived from subjects' responses typically predict the criterion values better than do the judgments on which the models were based (Goldberg, 1973). Dawes (1971) called this effect "bootstrapping." Such models are particularly effective when there is error in the judgments and when the attributes are positively intercorrelated. Indeed, Dawes and Corrigan (1974) and Wainer (1976) have shown that in such situations even equal-weight models predict the criterion well. Equal-weight models disregard all aspects of judgment except the judgment of whether the criterion tends to increase or decrease with an increase in each attribute. Given this judgment of sign, the equal-weight model is a simple summation of the standardized attribute values.

The predictive accuracy of linear additive models depends in part upon the options or items in the judgment set. The condition of positive intercorrelations among attributes is most likely to be met across a large heterogeneous set of options (for example, in general, better-furnished apartments are found in nicer neighborhoods). Thus, the model will do a good job of ordering such options. However, it will be less helpful in identifying the best option from the subset of top contenders. The set of serious contenders is the subset of options that lie on the <u>Pareto frontier</u>. This is the subset of options that are not dominated by some other option. An option is dominated if at least one other option is at least as good as it on all attribute dimensions. By definition, the options on the Pareto Frontier are better than one another in some respects and worse in others, so the attributes are negatively correlated across the subset (even if they are highly correlated across the entire domain). Such negative correlations weaken the predictive power of linear additive models (Stillwell, Seaver & Edwards, 1981).

Validation of both axiomatized and non-axiomatized value models has also been attempted in situations in which no criterion value is known. Of the several such approaches to validation, the most unsatisfactory is to compare the output of a decomposition model with unmodeled wholistic value judgments (e.g., Fischer, 1977; Huber, Daneshgar & Ford, 1971). High correlations may mask important deviations from the model or reflect the use of a very heterogeneous set of options for which all plausible procedures will produce similar orderings of the options. Furthermore, whenever people have difficulty integrating many items of information, wholistic judgments are an error-prone, flawed criterion. Thus, methods yielding low correlations with wholistic evaluations may be preferable to those that correlate highly with them.

A more sophisticated procedure is convergent validation, used by Fischer (1975) in comparing two wholistic methods and two methods based on additive value models in preference judgments for compact cars described by 3 or 9 attributes. The overall values produced by the two additive value models were in greater agreement than were those for the two wholistic methods. Eckenrode (1965), using experts, showed high convergent validity across six weight estimation procedures.

Short of entirely rejecting wholistic judgments, some researchers have argued that a small number of paired-comparison wholistic choices might be a relatively trustworthy expression of people's values. Thus, Schoemaker and Waid (1982) used 20 such binary choices as a criterion for evaluating 14 different value models. The task involved judging the suitability for college admission of high school graduates described on four attributes. The predictive accuracy of the different models did not vary much, except that linear value functions performed better than non-linear functions and unit weighting did poorly. MacCrimmon and Siu (1974) had subjects generate two indifference curves, one using paired-comparison choices and the other based on equivalence judgments like the following:

> ...The current [fiscal] policy [of an unspecified country] leads to an inflation rate of 8% and an unemployment rate of 8%. At what new inflation rate, paired with a new unemployment rate of 6%, would you feel indifferent toward the new policy and the one currently in effect? (p. 683)

The resultant curves were never identical, although in some cases they were very close. After both curves were elicited, a computer generated one critical pair of stimuli that would lead to different choices under the two indifference curves. In all cases, the choice made for the critical pair was consistent with the tradeoff curve generated by the paired-comparison choices.

One way of circumventing the privateness of people's values is by teaching subjects artificially designed "real values." These values can then be compared with those derived from different assessment methods. Using the multiple-cue probability learning paradigm, John and Edwards (1978) taught their subjects the worth of diamonds described by four attributes (cut, color, clarity, and carat). They found that the values derived from several weight-estimation methods closely matched the "true" values and that most subjects' weights provided a great improvement over equal weights. Whether these results can be generalized to situations in which values are private depends, in part, upon how explicit instruction affects values. Stillwell, Barron, and Edwards (1983) asked experienced bankers to use several different decomposition methods for evaluating the acceptability of loan applications.

Several of the methods produced evaluations similar to those produced by the bankers' usual computational procedure.

Noting the success of simple models in prediction situations and the difficulty people experience in performing complicated mental computations, Dawes (1977) proposed that these constitute theoretical grounds for trusting values produced by decomposition procedures more than values produced by wholistic judgment. Assuming that people can think clearly about what attributes should be important to them and how well each option rates on each attribute, then the model can hardly help but produce better summaries than the individuals themselves.

Riskless Value Models: An Assessment

Much of the interest in value models has centered around applications. Keeney and Raiffa (1976) wanted to help decision makers who were willing to spend the time and energy to think through their complex problems thoroughly. Edwards (1977) wanted an easy decision-making tool that could be widely used. Hammond and his colleagues focused on methods for resolving policy conflicts. Therefore, the literature has emphasized the development of decision aids rather than the descriptive adequacy of the models. Indeed, the justification for decision aids is that people do not usually make good decisions.

The strength of riskless value models rests on their ability to capture simpler component values and combine them to identify the overall evaluations that people would have if they had unlimited, errorless computational capacity and time for rumination. But what if people do not have consistent values or do not have any directly relevant values at all until asked? Then the method of asking may shape or create the values. A further concern is that the elicitation methods prescribed by the models are vulnerable to the potential sources of bias identified in the psychophysical literature, such as anchoring, recency, range, and other contextual effects (Poulton, 1982).

Multi-Attribute Utility Models

We return to risky utility theories, examining them now in their multiattribute form. Either the von Neumann-Morgenstern or the Savage formulation can be expanded to treat multi-attribute risky situations. Two general approaches may be taken (von Winterfeldt, 1975). Under the first, one starts with a riskless value model, $v(a) = f(v_i(a_i))$, where i indexes the n attributes of an object, a. Then a function h is constructed that transforms the overall value function, v, into a risky utility function, u:

$$u(a) = h(v(a)).$$

The methods used to measure $v(a)$ depend upon the particular value model used, but the stimuli presented to the assessor will always be riskless. For example, one might use dual standard sequences ("How many more miles would you be willing to drive to work for a rent decrease of...?"). Only in the final step, that of transforming the overall value function into an overall utility function, is uncertainty introduced. The transformation function h is found by using standard gambles, that is, finding indifference between a sure thing w and a gamble (x,p,y), where the outcomes w, x, and y are all multiattributed objects.

The second general approach is to assess the utilities directly, bypassing value functions (Keeney & Raiffa, 1976). A variety of models are available, each specifying a different combination rule by which the utilities of the attributes, $u_i(a_i)$, are combined into an overall utility for the object, $u(a)$. Here all the elicitation procedures involve gambles.

Among multiattribute utility models, the stronger the assumptions, the simpler the elicitation methods. Attributes are <u>mutually utility independent</u> if preference orderings over two gambles that vary in one or more attributes but are constant for the remaining attributes do not change as a function of the particular levels of the attributes held constant. For example, consider possible job offers varying in salary, location, and commuting time. Suppose you would prefer a sure job paying $25,000 in San Francisco with a 10-minute commuting time ($25,000, S.F., 10) to a 50-50 gamble yielding either ($15,000, S.F., 10) or ($35,000, S. F., 10). Then to satisfy utility independence for the attribute salary, you should also prefer a sure job of ($25,000, Denver, 60) over a 50-50 chance of either ($15,000, Denver, 60) or ($35,000, Denver, 60). Mutual utility independence requires this sort of consistency for each attribute and all subsets of attributes.

If mutual utility independence holds, then either the additive or the multiplicative model is satisfied. In the multiplicative model, the overall utility is the weighted sum of all the single attribute utilities and all possible cross-products of attribute utilities. A bit of algebra simplifies this to the following form:

$$1 + ku(a) = \prod_{i=1}^{n} (1 + kk_i u_i(a_i)).$$

The overall utility and all the attribute utilities are scaled from 0 (the worst level) to 1 (the best level). The k_i and k are scaling parameters; the k_i cannot be interpreted as "importance weights" because they depend on the relative ranges of the attributes. Thus, for example, if all job offers had approximately the same salary, k_i for salary would be small, even if, in general, money is very important to the decision maker. The other parameter, k, is a measure of multi-attribute risk aversion (when $k < 1$) or risk proneness (when $k > 1$). It is a function of the other scaling parameters, k_i, and need not be assessed directly.

The additive model,

$$u(a) = \sum_{i=1}^{n} k_i u_i(a_i),$$

is a special case of the multiplicative model and requires an additional assumption called <u>additive independence</u>: Attributes are additive independent if preferences over gambles depend only on the marginal probability distributions of the attributes and not on their joint probability distribution. One test for additive independence is that the k_i sum to 1. Another test is that you should be indifferent between two gambles, one of which offers a 50-50 chance of getting either (a) the best object or (b) the worst object and the other of which offers a 50-50 chance of getting either (c) an object for which some attributes are at their best level and the rest are at their worst level or (d) an object for which those attributes that were at their best level in (c) are here at their worst level and vice versa. For example, (a) might be your favorite phonograph record and \$10; (b) your least favorite record and \$.10; (c) your favorite record and \$.10; (d) your least favorite record and \$10.

For either the multiplicative or the additive model, each single-attribute utility function can be elicited using standard gambles, holding constant (at any convenient level) the levels of all but one attribute. The scaling parameters, k_i, may be found by eliciting, for each attribute in turn, a probability, p_i, that makes you indifferent between (a) receiving the object having the best level of attribute i and the worst level of all other attributes versus (b) a gamble that pays the best object with probability p_i and the worst object with probability $(1-p_i)$. Then $p_i = k_i$. When the so-called "corner" objects, all of whose attribute levels are either the worst or the best, seem unrealistic, other, more lengthy, elicitation techniques may be used.

Fischer (1977) devised a cover story to facilitate use of the standard gamble techniques. The stimuli were multi-attribute job descriptions. Subjects identified the worst job, J_*, and the best job, J^*. They were told to imagine that they had been firmly offered J_* and that they might also be offered J^* some time from now. For each intermediate job, J', they were to imagine that they had just received a job offer, which they must accept or reject immediately:

> Clearly your decision in this matter will depend upon how likely you think you are to receive the J^* offer. Your task...is to specify for each offer J' a probability p of receiving the J^* offer such that you would be indifferent between accepting or rejecting the J' offer (p. 308).

These multi-attribute risky models are intended for practical application. However, they are so complicated that they require an extended interaction between the decision maker and an analyst. The decision maker should be an expert in the problem area; the analyst, an expert in the methodology. Keeney (1977) has published his elicitation of an energy expert's utility model and functions over 11 attributes, including deaths, pollution, nuclear safeguards, and electricity generated. The model was designed for evaluating energy policies differing in type of fuel (fossil or nuclear) and degree of conservation. The dialogue between Keeney and the energy expert shows the intense cognitive effort required of both individuals.

INFORMATION PROCESSING IN DECISION MAKING

The experimental study of decision making has paralleled the emergence of the field of cognitive psychology. This field, with its emphasis on internal processes, mental limitations, and the way in which the processing of information is shaped by these limitations, has come to have a profound influence on decision theory and research. Examples of this influence include prospect theory, with its emphasis on the decision frame, and the EBA model, which is essentially a strategy for reducing the strain of choosing among many complex alternatives.

This section examines the information-processing approach to decision making. Although we shall continue to distinguish between risky and riskless choice studies, this distinction is less relevant than in the theories of choice described earlier. The information-processing paradigm represents alternatives as multidimensional stimuli, with outcome probability as just another, albeit somewhat special, dimension. Many of the same cognitive processes can be observed in both risky and riskless settings.

Confronting Human Limitations: Bounded Rationality

The traditional view of human beings' higher mental processes assumes that we are, in Shakespeare's words, "noble in reason, infinite in faculties." A 20th century expression of this esteem was provided by a well-known economist who asserted, "we are so built that what seems reasonable to us is likely to be confirmed by experience or we could not live in the world at all" (Knight, 1921, p. 227).

Research in cognitive psychology has painted a much more modest picture of human capabilities. In his influential study of classification and coding, Miller (1956) demonstrated the limitations of people's ability to attend to and process sensory signals. About the same time, Bruner, Goodnow and Austin (1956) concluded that subjects in their concept formation tasks were experiencing a condition of cognitive strain, which they attempted to reduce by simplification strategies.

In the study of decision making, too, the classic view of behavioral adequacy was being challenged on psychological grounds. A leading critic of utility maximization was Simon (1959), who observed:

> The classical theory is a theory of a man choosing among fixed and known alternatives, to each of which is attached known consequences. But when perception and cognition intervene between the decision-maker and his objective environment, this model no longer proves adequate. We need a description of the choice process that recognizes that alternatives are not given but must be sought; and a description that takes into account the arduous task of determining what consequences will follow on each alternative (p. 272).

As an alternative to the maximization hypothesis, Simon introduced the notion of <u>bounded rationality</u>, which asserts that cognitive limitations force decision makers to construct simplified models of their problems. Simon argued that the decision maker behaves rationally with respect to this simplified model. To predict decisions, we must understand how this simplified model is constructed through processes of perception, thinking, and learning.

According to Simon, the key to simplification was the replacement of the maximization goal by the <u>satisficing</u> principle: outcomes are first classified as "satisfactory" or "unsatisfactory" with respect to each of the relevant attributes; the first alternative that satisfies this level of aspiration for each attribute is selected. In evaluating investment plans, for example, one may select the first plan that provides satisfactory profit as well as adequate security. What is considered to be a satisfactory profit may change with time and experience as one's aspiration level increases or decreases.

Satisficing is simpler than utility maximization in several important respects. It bypasses the problems of evaluating the overall utility of each outcome or of comparing diverse attributes. It does not call for detailed exploration of all the available alternatives and it requires only a very limited computational capacity.

Like other models of decision making, bounded rationality has predictive power only if its primitives can be specified. That is, one must be able to tell, independently of people's choices, what attributes they consider and what their levels of aspiration are on each. Information-processing studies have been used to provide such evidence (see, e.g., Payne, Laughhunn & Crum, 1980).

Methods for Studying Information Processing

There are three major categories of experimental methods used to provide insight into decision processes: (a) inference from stimulus/response studies, (b) algebraic modeling of wholistic judgments, and (c) process-tracing methods.

Inference studies. Inferences about mental strategies can often be derived from responses made to specifically constructed stimuli. Shaklee and Tucker's (1980) investigation of the strategies used in assessing degree of relatedness follows this pattern as does the use of duplex gambles to determine the relative attention paid to surface characteristics and underlying probability distributions (Payne and Braunstein, 1971; Slovic and Lichtenstein, 1968a). Another example of this approach is a study by Huber (1983), who constructed pairs of alternatives, x and y, such that the application of one choice rule led to selection of x, the application of a second rule led to selection of y, and use of any other rule (from a set of five possibilities) led to no decision. Choices among a set of such pairs ordered the rules in terms of popularity.

Algebraic models. The lens model, previously discussed, is one of several algebraic models used to study how decision makers weigh and combine information from multiple sources (Slovic & Lichtenstein, 1971). The multiple regression analysis used in the lens model produces an equation:

$$\hat{Y} = \sum_{i=1}^{n} b_i X_i$$

where $\hat{Y}$ is the predicted judgment and the b_i are interpreted as the weights given to the cue dimensions, X_i. The equation can be expanded by adding exponential terms to model nonlinear use of the cues or cross-product terms to model non-additive combination rules (Einhorn, 1970).

Evidence about information processing has been obtained by studying changes in lens model measures as the task or stimulus set is changed. For example, Hoffman (1968) reported data showing that the relative weights of two cues changed as a function of cue consistency. For stimulus sets in which the two cues tended to have similar values (the cues were congruent in their implications for the criterion value), the judges weighted the two cues approximately equally. However, in response to stimulus sets in which the cues were incongruent, judges weighed one cue more heavily than the other.

Whereas the lens model assumes that the decision maker operates on the cue dimensions as given, Anderson (1970, 1974, 1981) has developed methods for simultaneously scaling the subjective stimulus values and determining the weighting parameters. The resulting models are used to test theories about the information integration rules used by subjects. Particular attention has been given to tasks in which a simple algebraic model, involving adding, averaging, substracting, or multiplying the informational inputs, serves as the substantive theory of judgment that is being tested.

Anderson's approach, called information integration theory, uses stimuli created by factorial combinations of information dimensions. The subjects' wholistic judgments are analyzed with an analysis-of-variance model. The model incorporates both a theory (the form of the model) and a goodness-of-fit test for the data. An invalid response scale could cause a valid model to fail the test of fit. Therefore, Anderson's approach performs a monotone rescaling of the response variable. Failure to find any rescaling that will make the data fit some version of the basic model argues against the model, and success argues for it. Once the model and response scale are established, the subjective values of the stimuli can be derived. Anderson uses the term functional measurement to describe this interplay between theory and scaling.

Another method that assumes the form of a model and then attempts to fit data to it is conjoint measurement (Debreu, 1960b; Krantz & Tversky, 1971; Luce & Tukey, 1964; Tversky, 1967b). Here, the ordinal properties of the judgments are used to test proposed rules for integrating items of information. The simple algebraic theories tested by functional measurement and conjoint measurement methods have been found to provide excellent fits to data in such diverse domains as social cognition, developmental psychology, psychophysics, and linguistics, as well as decision making.

Despite their successes, these algebraic modeling techniques have serious limitations. First, they require many judgments. Judges facing a long, perhaps boring, task may resort to simplifying strategies that do not represent their usual ways of thinking (Slovic, Lichtenstein & Edwards, 1965). In order to mitigate this problem, Barron and Person (1979) have proposed a method (called HOPE), based on a highly fractionated orthogonal design, that requires many fewer judgments.

Additional problems arise from the deficiencies of representative and orthogonal designs. Orthogonal designs may include stimulus profiles that are so peculiar and so unlikely to occur that the judge cannot reasonably evaluate them. In representative designs, the stimulus dimensions are typically intercorrelated. As a result, the derived weights are non-unique and difficult, if not impossible, to interpret (Darlington, 1968).

Finally, an algebraic model's ability to model a set of responses is no guarantee that it captures the psychological processes that produced those responses. As Hoffman (1960, 1968) noted, two or more models may be algebraically different yet equally predictive, given fallible data. Furthermore, two or more models may be algebraically equivalent yet suggest radically different processes. Drawing an analogy to problems of classification in minerology, Hoffman introduced the term <u>paramorphic representation</u> to remind researchers that "the mathematical description of judgment is inevitably incomplete ... and it is not known how completely or how accurately the underlying process has been represented" (1960, p. 125).

Lopes (1982c) recognized the need to go beyond mathematical equations to understand the fine structure of the cognitive mechanisms involved when people produce averages, products, or other algebraic forms. Noting that judges are typically not conscious of these computations or of the equations they represent, she asked: "What psychological processes give rise to this algebra-less algebra?" (p. 1). To answer this question, she proposed a procedural theory in which judgments are produced by a serial process whereby an initial quantity or anchor is adjusted one or more times in accordance with other available information. Lopes argued that this process can plausibly account for a wide variety of judgments that have been modeled by algebraic functions. A thorough test of Lopes' theory requires use of special experimental methods, such as those that we describe next.

Process-tracing methods. In contrast to methods that make inferences about unseen processes intervening between stimulus and response, process-tracing techniques (Raaij, 1983, Svenson, 1979) attempt to make these processing strategies directly observable. There are three main process-tracing methods: verbal protocols, information monitoring, and eye-movement analysis. Response time analysis has also been used, but to a lesser extent.

Verbal protocols require subjects to think out loud as they perform their decision tasks. This approach differs from the introspective methods employed in the early days of experimental psychology (Tichener, 1910). Introspection used retrospective reports by highly trained subjects, whereas verbal protocols attempt to capture thoughts as they occur, using subjects who are relatively naive about the researcher's theories. For example, subjects might be instructed to report what information they consider as they examine an alternative, to describe each thought they have about that alternative, and to verbalize the reasoning that leads them from observation to decision. Typically, the protocol is partitioned into short phrases corresponding to single cognitive operations. These are then coded and analyzed to test or design process models (Newell & Simon, 1972).

Several concerns have been raised regarding the validity of verbal protocols. One is that subjects cannot report accurately on their own mental processes. Nisbett and Wilson (1977) reviewed the literature and concluded that subjects tend to describe what they believe their mental states should have been, not what they actually were. A second concern is that the act of reporting and the instructions regarding what to report may distort the processes (Flaherty, 1975; Lichtenstein, 1982; Posner, 1982). In addition, people may not be able to articulate all their internal states (Lindsay & Norman, 1972). Some, in fact, hold that this is a hallmark of substantive expertise (Polanyi, 1958). Ericsson and Simon (1980) have rebutted many of these criticisms, noting, in particular, that many of Nisbett and Wilson's observations pertained to retrospective rather than concurrent protocols. In defending verbal protocols, Hayes concluded:

> Analyzing a protocol is like following the tracks of a porpoise. Occasionally, the porpoise reveals itself by breaking the surface of the sea. Its brief surfacings are like glimpses which the protocol affords us of underlying mental process. Between surfacings, the mental process, like the porpoise, runs deep and silent. Our task is to infer the course of the process from these brief traces" (1982, p. 77).

An eye-movement protocol records where the subjects fix their gaze as they perform a decision task (Russo & Rosen, 1975; Russo & Dosher, 1975). The sequence of fixations produces a detailed trace that may be harder for subjects to censor and better suited for rapid processes than are verbal protocols. On the negative side, the measurement apparatus is quite obtrusive and often restricts the stimuli to simplistic displays. To obtain an accurate record, the subject's head must be immobilized and the items widely separated. In addition, eye fixations reflect information-seeking responses and, hence, cannot reveal all details of internal processing.

For both verbal protocols and eye-movement protocols, data collection is time consuming, producing masses of data requiring detailed analysis. Information-monitoring methods restrict the data to a simpler type, suitable for testing hypotheses regarding information search (Jacoby, 1975, 1977; Payne, 1976). The typical study presents information on a display board, a matrix array with alternatives as rows and attributes as columns. Information is available in each cell of the matrix, giving the value for the particular attribute and alternative. Subjects choose an option after selecting as much information as desired. The sequence of selections can reveal, for example, whether people first examine the value of all alternatives on a given attribute (as might be predicted by the EBA model) or try to get a fuller picture of individual alternatives. The disadvantages of this method are its obtrusiveness, its inability to provide insight regarding the use of information stored in memory, and its lack of informativeness regarding how acquired information is processed.

Results from any of these process-tracing methods can be represented in terms a <u>flow diagram</u> (e.g., Newell, 1980) or <u>decision net</u> (Bettman, 1979). An early example was produced by Clarkson (1962), who attempted to simulate trust investment officer's portfolio selection process (see Figure 13). To evaluate his model, Clarkson fed the relevant information about various stocks into a computer model analogous to the net. The model matched 24 of the officer's 29 stock selections. Decision nets have been used to model medical and psychiatric diagnoses (Kleinmuntz, 1963), accounting decisions (Bouwman, 1982), and consumer product choices (Bettman, 1979). Bettman (1979) described a variety of methods for analyzing and characterizing the structure, reliability, and efficiency of these nets.

Insert Figure 13 about here

Integration of methods. Advocates of process tracing have expressed serious reservations about the ability of algebraic models to represent mental operations. For example, Simon (1976), stated that: "The variance analysis paradigm...is largely useless for discovering and testing process models to explain what goes on between the appearance of stimulus and performance of response" (p. 261). Einhorn, Kleinmuntz, and Kleinmuntz (1979) defended the algebraic models by arguing, theoretically and empirically, that both process tracing and algebraic methods can provide valid descriptions of the same processes, although at different levels of detail. Further, they advocated employing the two techniques together in a multi-method approach to take advantage of their complementary strengths and offset their respective weaknesses. A similar recommendation was been put forth by Payne, Braunstein, and Carroll (1978). The procedural theory of judgment proposed by Lopes (1982c) provides a nice illustration of how process training studies might usefully complement algebraic modeling.

Information-Processing Findings

The methods described above have been used by researchers with two general objectives in mind. The first is to discover the elementary operations and rules that are employed in decision making. The second is to determine how features of the decision task govern the selection and use of these rules.

Basic rules. The decision rules that have been observed can be categorized in a number of ways. Some, like the linear additive model, are compensatory, meaning that a high score on one dimension can offset a low score on another dimension. In contrast, non-compensatory processes do not permit trade-offs among dimensions. For example, a conjunctive rule eliminates any alternative that fails to surpass a criterion value on any dimension. It is the basic rule of satisficing. A disjunctive rule selects any alternative that surpasses the criterion for at least one dimension. A lexicographic process chooses the alternative that is superior on the most important dimension. If more than one alternative has top ratings on this dimension, then the next most important dimension is considered, and so forth. Tversky's EBA model is a probabilistic version of this rule.

Another way of categorizing an information-processing pattern is according to whether it focuses on alternatives or on attributes. Intra-alternative rules consider all the attributes of each alternative before going on to the next alternative. Dimensional rules compare all alternatives on one dimension at a time. Compensatory and conjunctive rules are instances of intra-alternative processing. The lexicographic and disjunctive rules involve dimensional processing.

Effects of task difficulty and context. Pure reliance on any one strategy is uncommon in information-search tasks; subjects typically alternate between intra-alternative and dimensional processing (Bettman & Jacoby, 1976). In decision-making tasks, the strategy used seems to depend upon task characteristics such as difficulty, context, and familiarity. Task difficulty has been studied by varying the number of alternatives, the number and relevance of dimensions or attributes, and the information display.

Svenson (1979) manipulated difficulty by varying the number of attributes (from 2 to 12) and number of alternatives (from 2 to 13). His subjects examined less information as both alternatives and attributes increased, with increases in the number of attributes having the greater effect. As the number of alternatives increased, there was a shift from dimensional to intra-alternative processing along with increased use of absolute statements (e.g., "this apartment is big") relative to comparative statements (e.g., "that apartment is bigger").

Wright (1974a, b) found that making the task more difficult by including incomplete or irrelevant data, noncomparable scaling across dimensions, and time pressure increased the use of elimination strategies such as the conjunctive and lexicographic rules. Because negative (unfavorable) data are conducive to elimination strategies, such data tend to be used more as task difficulty increases. Gaeth and Shanteau (in press) demonstrated that presence of irrelevant information impaired performance in an applied setting (soil judgment). However, this impairment was reduced by training.

Tyszka (1983) observed strong effects of context in choices among multiattribute alternatives. Introduction of an alternative C, which was dominated by alternative B, enhanced B's chances to be chosen over a third alternative, A, in violation of the principle of independence from irrelevant alternatives.

Another contextual effect is that dimensional processing generally occurs more often when the alternatives have dimensions or attributes in common (Capon & Burke, 1977; Russ, 1971; Russo & Dosher, 1975). Indeed, the attractiveness of performing dimensional comparisons can give an otherwise secondary dimension greater importance if all alternatives are characterized on it (Slovic & MacPhillamy, 1974). Tversky (1969) noted that processing by dimensions is often more efficient, in the sense of requiring fewer operations, allowing small differences to be ignored, and facilitating the use of the dominance rule to eliminate inferior alternatives.

Effects of display. Numerous studies have found that decisions can be sensitive to how information is displayed. Friedman (1966) and Branscombe (1975) found that few consumers could perform the mental calculations needed to select the most economical products in the marketplace. Unit pricing is a remedy for this problem. However, Russo (1977) found that consumers failed to use unit prices when they were displayed under each brand. Only if the unit prices were displayed in a simple organized list was this information used. Huber (1980) showed that decision processes were influenced by whether information was described in numerical rather than verbal form. Numerical presentation led to more dimensional comparisons and fewer comparisons against a criterion.

Display effects highlight the difficulties that decision makers have in making tradeoffs and performing even simple mental calculations. Such difficulties led Slovic (1972) to observe "that a judge or decision maker tends to use only the information that is explicitly displayed in the stimulus object and will use it only in the form in which it is displayed. Information that has to be stored in memory, inferred from the explicit display, or transformed tends to be discounted or ignored" (p. 14).

<u>Phased strategies</u>. Many studies have found evidence for phased strategies, in which an initial set of rules, designed to eliminate alternatives, are followed by more detailed evaluation of the surviving options (Wright & Barbour, 1977). Phased strategies seem particularly likely when the number of alternatives is large (Sheridan, Richards & Slocum, 1975; Svenson, 1974). A verbal protocol illustrating a multiple-rule sequence is shown in Table 5. The subject used an elimination-by-aspects process to reduce the choice problem from 12 to eight, and eventually to just two alternatives. At that point, an additive-difference strategy was used.

Insert Table 5 about here

Although intuitively appealing, the application of an initial screening phase can lead to suboptimal decisions. Removal of dominated alternatives is certainly defensible. However, other elimination strategies may lead to rejection of options that would, under more thorough scrutiny, prove to be quite attractive.

<u>Selection of decision rules</u>. Payne (1982) has offered an account of how people choose decision rules. He points to three rule-selection processes: cost/benefit principles, perceptual processes, and adaptive production systems. Cost/benefit principles involve a trade-off between optimality and simplicity in choosing a way to process information. They are similar to the economists' notion of "transaction costs," which determines how hard people will work on making decisions and can be used to explain away apparent instances of suboptimality (by saying that it was not worth the added effort to identify the very best alternative). The most elaborate model of this sort

is that of Beach and Mitchell (1978; see also Christensen-Szalanski, 1978). In order to have any empirical content, such models need an independent measure of mental effort. Shugan (1980) proposed measuring the cost of thinking in terms of the number of single-attribute comparisons needed to discriminate between two alternatives. Johnson (1979) developed a measure of effort based on the number of elementary operations (e.g., multiplication, addition, subtraction, and comparison) required to apply a rule in a given situation.

Perceptual processes are involved when the selection of decision rules is done in a non-deliberative manner. For example, Tversky and Kahneman (1981) proposed that decision frames shape the choice of decision rules in as automatic a way as different spatial perspectives shape visual perceptions. In contrast to the cost/benefit theory, the perceptual analogy implies that subjects will typically be unaware of the effects of alternative frames, consistent with the findings of Fischhoff (1983).

The production-systems approach was adapted by Pitz (1977) from information-processing theory (Newell & Simon, 1972). A production system is a condition-action pairing such as: "If you have the values of two alternatives on the same attribute, then compare the values and note which alternative is better." Newell (1980, p. 704) called such systems "a species of pattern-directed rule-oriented program system" that are active candidates "for the underlying architecture of human cognition."

Justification and Choice

One appealing aspect of relying on a series of deliberative rules is that it produces a set of reasons for justifying the alternative that is eventually selected. Tversky (1972) invoked justifiability as a reason why people might use elimination by aspects. That process leads to a clear-cut choice without recourse to relative weights, trade-off functions, or other numerical computations that are hard to describe to those not versed in decision theory.

The importance of justification processes can be seen in a study of difficult choices (Slovic, 1975). Each of two options was defined by two dimensions differing in importance. To maximize the difficulty of choice, the paired options were designed to have equal worth by making the option that was superior on the more important dimension so inferior on the lesser dimension that its advantage was cancelled. For example, one pair of options involved gift packages with two components, cash and a coupon book offering miscellaneous products and services with a stated monetary value. The subject was shown two such gift packages with one component missing, for example:

	Cash	Coupon Book Worth
Gift package A	\$10	----
Gift package B	\$20	\$18

The subject supplied a value for the missing component such that the two options would be equally attractive. After equating various pairs of options, subjects made choices from the equated pairs. Contrary to most choice theories, decisions regarding these equally attractive alternatives were not made randomly. Rather, most subjects consistently selected the option that was superior on the more important dimension. Apparently, reliance on the more important dimension makes a better justification ("I chose this gift package because it provided more cash") than random selection ("They looked about equally attractive, so I flipped a coin").

Tyszka (1981) and Montgomery (1983) have both advocated theories based on the concept of justification. Montgomery proposed a model that describes the decision process as a search for a dominance structure, whereby decision makers restructure decision problems until they find a perspective that shows a (relatively) conflict-free way to make a choice. This search may involve bolstering or deemphasizing the importance of certain attributes or collapsing two or more attributes into a more comprehensive one.

Script Processing

Abelson (1976) has criticized traditional theories of decision making as being "overly elementaristic, stilted, and static" (p. 33). He proposed an alternative theory, based on the concept of a script, defined as a coherent sequence of events expected by an individual. Scripts are learned throughout one's lifetime, by experiencing and observing event sequences. A restaurant script, for example, involves a set of expectations about food prepared and served, about menus, waiters, tips, and checks.

Abelson (1981) noted that scripts can provide guides or strategies for decision making. Abelson (1976, p. 37) described several scripts that might influence decisions about graduate school applicants made by a member of an admissions committee. For example, in an episodic script, a past single case would be recalled, similar to the applicant under consideration: "Mr. X reminds me very much of Mr. Y, who hung around for eight years never writing his dissertation. Let's not get into that again." A categorical script assimilates the applicant to a category: "He's one of those guys who writes about all this existential stuff and ends up wanting to go into clinical psych." Although many behaviors observed in studies of decision making can be interpreted in terms of scripts, Abelson admits that the theory needs much more precise specification.

Information Processing in Risky Choice

Gambles as multidimensional stimuli. Slovic and Lichtenstein (1968b) proposed that gambles could be characterized in terms of four basic dimensions--probability of winning (P_W), amount to win ($\$_W$), probability of losing (P_L) and amount to lose ($\$_L$). They argued that an adequate descriptive theory of risk taking required an understanding of how people integrated these dimensions when evaluating gambles.

In attempting to understand the relative influence of these risk dimensions, Slovic and Lichtenstein hypothesized that an individual assigns weight to a particular risk dimension according to its perceived importance. However, information-processing limitations may restrict the individual's ability to act on the basis of those importance beliefs when evaluating a gamble.

To test their hypotheses, Slovic and Lichtenstein (1968b) presented several groups of subjects with duplex gambles (see Figure 5) that allowed all four risk dimensions to vary independently. Several groups of subjects expressed their attitudes toward playing these gambles, but did so using different response modes. One group rated the attractiveness of each gamble, whereas each of three other groups used a bidding method in which the gamble was equated with an amount of money (either the maximum buying price, the minimum selling price, or the monetary equivalent). A regression analysis indicated large differences in the relative weights of the various risk dimensions. On the average, the highest correlation between a subject's evaluations and one of the four risk dimensions was twice the size of the lowest correlation. Attractiveness ratings correlated more highly with P_W than with any other dimension, whereas the bidding responses correlated most highly with the payoff dimensions, particularly $\$_L$. Slovic and Lichtenstein (1968b) suggested that the reference to money in a bid focussed attention on the monetary aspects of the gamble, whereas the rating focussed attention on the probabilities.

Although not designed specifically to test the SEU model, these results cast doubt on its adequacy. According to the model, all risk dimensions should have equal influence and evaluations of gambles should be invariant across response modes. The Slovic and Lichtenstein results are, however, compatible with an information-processing perspective.

Integration rules. Slovic and Lichtenstein showed that an additive model, with each risk dimension weighted by its correlation with the responses, fit thier data well (Slovic & Lichtenstein, 1968b). However, Anderson and Shanteau (1970) found a better fit with a modified SEU model:

$$R = W_W S_W + W_L S_L$$

where S_W and S_L corresponded to subjective versions of $\$_W$ and $\$_L$ and the W_W and W_L were subjective functions of the probabilities of winning and losing. Unlike subjective probabilities, however, these weights do not have to sum to 1.0.[5]

Payne (1980; Payne & Braunstein, 1978) elaborated the Anderson and Shanteau model by proposing that it implied (a) intra-alternative processing of alternatives, involving many probability-amount comparisons (as opposed to probability-probability or amount-amount comparisons); and (b) a compensatory process in which equal amounts of information are sought for each alternative. He noted that these implications were not confirmed by the results of several process-tracing studies. These studies showed that the amount information searched varied from one gamble to another. Furthermore, this variation increased with the number of gambles in the choice set, many gambles being eliminated after only a limited amount of search. As the number of gambles increased, the proportion of available information searched decreased. There was more intra-dimensional search as the number of gambles increased. Payne concluded that, with more gambles to consider, subjects are less likely to use strategies consistent with a compensatory algebraic model. Payne has proposed a contingent, multi-stage model to account for these and other choice data.

Response mode effects. Individuals can express their preferences in many different ways. Although the decision maker is sometimes free to determine the mode of response, more often some external source defines it. A common (and critical) distinction is whether the task is presented as one of judgment (evaluating individual options) or as one of choice among two or more options. Most theories of decision making view judgment and choice as equivalent. However, numerous empirical studies have found that information-processing strategies used in making choices are quite different from the strategies employed in judging single alternatives. As a result, choices and evaluative judgments of the same options often differ, sometimes dramatically.

An early demonstration of response-mode effects in risky choice was the finding by Slovic and Lichtenstein (1968b) that ratings of a gamble's attractiveness and choices among pairs of gambles were influenced primarily by the probabilities of winning and losing, whereas buying and selling prices were primarily determined by the dollar amounts that could be won or lost. When subjects found a bet attractive, their prices correlated predominantly with the amount to win; when they disliked a bet, their prices correlated primarily with the amount that could be lost. This pattern of correlations was explained as the result of a starting point (anchoring) and adjustment procedure used when setting prices. Subjects setting a price on an attractive gamble appeared to start with the amount to win and adjust it downward to take into account the probability of winning, the possibility of losing and the amount that could be lost. However, the adjustment process was typically inadequate, leaving the price response unduly influenced by the starting point payoff. Choices, on the other hand, appeared to be governed by different rules, such as dimensional comparisons of the gambles' probabilities.

Lichtenstein and Slovic (1971) hypothesized that, if people process information differently when making choices and setting prices, it should be possible to construct pairs of gambles such that the same individual would choose one member of the pair but set a higher price on the other. They demonstrated this predicted effect in several studies, including one conducted on the floor of the Four Queens Casino in Las Vegas (Lichtenstein & Slovic, 1973). A typical pair of gambles in that study consisted of:

Bet A: 11/12 chance to win 12 chips
1/12 chance to lose 24 chips

Bet B: 2/12 chance to win 79 chips
10/12 chance to lose 5 chips

where each chip was worth 25 cents. Each subject first made a simple choice, A or B. Later, the subject indicated a minimum selling price for each bet. For this pair of gambles, Bets A and B were chosen about equally often, across subjects. However, Bet B received a higher selling price about 88% of the time. Of the subjects who choose Bet A, 87% gave a higher selling price to B, thus exhibiting a preference pattern inconsistent with almost every normative and descriptive theory of preference.

These response-mode induced reversals of preference have been replicated in numerous other studies (reviewed by Slovic & Lichtenstein, 1983). Of particular interest is a study performed by Grether and Plott (1979), two skeptical economists concerned about the challenge that preference reversals pose for theories of choice. They conducted a series of experiments "to discredit the psychologists' works as applied to economics" (p. 623). Their design was based on thirteen criticisms or explanations that would render the preference-reversal phenomenon irrelevant to economic theory, including the fact that the experimenters were psychologists, which might have led the subjects to behave peculiarly. Their manipulations included using special incentive systems to heighten motivation, controlling for income and order effects, testing the influence of strategic or bargaining biases, and having economists conduct the study. To their surprise, preference reversals remained much in evidence despite their careful attempts to eradicate them. Not all economists have been so skeptical. Arrow (1982) has pointed out a number of failures of utility theory in non-experimental contexts such as insurance, securities, and futures markets, which he feels are directly interpretable in terms of information-processing factors.

Implications for utility theory. Hershey, Kunreuther, and Schoemaker (1982) have applied the findings from information processing studies to the problem of measuring utility curves. They identified five choices that must be made in order to select a method for eliciting a utility curve. They then showed that each of these choices can affect the shape of the curve elicited, although none of these factors is incorporated into utility theory.

First, what response mode is used? Hershey et al. demonstrated that the use of the certainty-equivalence method for eliciting utilities, in which the subject specifies the value of a sure outcome, w, such that the subject is indifferent between w and a given gamble, (x,p,y), leads to more risk-prone utility functions than the use of the probability-equivalence method, in which x, y, and w are given and the subject specifies the value of p. Second, what values are used in the standard gamble and the sure outcome? Variations in the experimenters' choice of values for x, y, w, and p led to systematic changes in the proportion of risk-averse responses, changes that could not be explained by utility theory. Third, what is the domain of the standard gamble? Subjects showed significantly more risk aversion for gambles containing both wins and losses than for gambles offering no possibility of winning. Fourth, who gets the risk? Subjects who were offered the opportunity to receive a gamble to play were more risk averse than those who were told they were already in the risky situation and were asked if they wished to transfer the risk to someone else. Fifth, what is the decision context? Subjects were more risk averse when gambles were presented with an "insurance" frame than when presented with a "gamble" frame.

Taken together, these results, along with other, related findings previously discussed, show that risk attitudes as defined in utility theory are easily manipulated, perhaps indeterminate. Thus the very concept of a utility function is in doubt.

PREFERENCE MANAGEMENT

Applications of Decision Theories

The study of decision making is an applied science as well as a theoretical one. Its aim is to help individuals make better decisions in their personal lives and in their jobs as managers, physicians, policy makers, etc. The studies described in this chapter have been scrutinized by those concerned with improving the practice of decision making. For example, the field of behavioral accounting applies the information-processing approach to financial analysis (Ashton, 1982). Market researchers have adopted expected utility and probabilistic choice theories as models of consumer decision making (Bettman, 1979; Jacoby, 1975). Medical Decision Making, The New England Journal of Medicine, and related publications report numerous applications of utility theory to medical diagnosis and treatment (Krischer, 1980).

The dominant methodology for aiding decision making is a blend of systems analysis, operations research, and utility theory called decision analysis (Howard, 1968; Keeney, 1982; Raiffa, 1968). It assumes that decision makers wish to select actions with the highest expected utility, based on their preferences and beliefs. The tools it offers are methods for structuring the decision problem and eliciting the decision maker's subjective utilities and probabilities. For example, when outcomes have many components, multi-attribute techniques can help assess and integrate the various utilities. Decision analysis has been applied to such diverse problems as hurricane seeding (Howard, Matheson & North, 1972), selecting experiments for a Mars space mission (Matheson & Roths, 1967), coronary artery surgery (Pauker, 1976), cancer chemotherapy (McNeil, Weichselbaum & Pauker, 1981), and family planning (Beach, Townes, Campbell & Keating, 1976).

Psychological studies have had a substantial effect on decision-aiding methods. The heuristics and biases observed by Tversky and Kahneman (1974) and others have changed how decision analysts elicit probabilities (Spetzler & Staël von Holstein, 1975). The work by Gettys et al. (1980) points toward improved methods for structuring event hypotheses and decision alternatives. Psychological considerations are increasingly incorporated in the design of decision support and management information systems (Benbasat & Taylor, 1982; National Academy of Sciences, 1983).

In considering the influence of behavioral research on choice theories, March (1978, 1982) noted that theories of rational choice presume two improbably precise guesses about the future. One guess concerns the future consequences of current actions. The other guess concerns future preferences among those consequences. March (1978) argued that behavioral research has already shaped the rational theory's treatment of the first guess. For example, economic theories now place considerable emphasis on ideas of search, attention, and information costs. Aspiration levels and satisficing have been described as sensible in many settings. The future may see similar progress devoted to the second guess.

Labile Values

Regarding the second guess, March (1978) argued that the same limited cognitive capacity that affects information processing about facts also affects information processing about values: "Human beings have unstable, inconsistent, incompletely evoked, and imprecise goals at least in part because human abilities limit preference orderliness (p. 598)." March drew upon a rich and diverse array of observations to argue that, contrary to normative theory, preferences are neither absolute, stable, consistent, precise, nor unaffected by the choices they are presumed to control. The framing and response-mode effects described above represent a few pertinent examples.

Even when cognitive capacity is not strained, preferences may be labile because we do not really know what we want or how we will experience certain outcomes. When considering simple, familiar decision consequences, one's preferences may be well articulated. But the most interesting and important decisions, such as medical treatments, marriage, and career choice, tend to have novel, unfamiliar, and complex outcomes. In such circumstances our values may be incoherent, not sufficiently thought through (Fischhoff, Slovic & Lichtenstein, 1980). When we think about societal risks, for example, we may have contradictory values (e.g., a strong aversion to catastrophic losses of life, but an awareness that we are no more moved by a plane crash with 500 fatalities than one with 300). We may occupy different roles in life (parents, workers, children), each of which produces clear-cut but inconsistent values. We may vacillate between incompatible but strongly held positions (e.g. freedom of speech is inviolate, but it should be denied authoritarian movements). We may not even know how to begin thinking about some issues (e.g., the appropriate tradeoffs between the outcomes of surgery for cancer vs. the very different outcomes from radiation therapy). We may underestimate our ability to adapt to extremely good or extremely bad circumstances (Brickman, Coates & Janoff-Bulman, 1978; Cameron, Titus, Kostin & Kostin, 1973). Our views may change so much over time (say, as we near the hour of decision or of experiencing the consequences) that we become disoriented as to what we really think.

At times, it seems as though there are rival selves within the same individual, each vying for legitimacy. Schelling (1982) pointed to people who set alarm clocks but do not respond to them, who want to quit smoking but cannot. Noting that robot chess players can be programmed to play at different levels of skill, he asked whether analogous signals in humans might tune in and tune out particular qualities of memory, perceptual acuity, motivation, and value, thereby selecting the individual who is to act in a particular setting. Thaler and Shefrin (1981) modeled self-control as a balancing of the interests of a "doer" and "planner" within each individual (with the former acting like an id and the latter like a superego, both expressed in terms of Lagrangian multipliers). A striking empirical demonstration of multiple selves is provided by Christensen-Szalanski (in press), who recorded the changes in attitudes of pregnant women toward anesthesia before, during, and after labor.

Fischhoff, Slovic, and Lichtenstein (1980) noted the problems that labile preferences pose for the measurement of values. Although some practitioners have been sensitive to the possibility that complex elicitation methods may induce errors of assessment (e.g., Bursztajn & Hamm, 1982; Edwards, 1977; Llewellyn-Thomas, Sutherland, Tibshirani, Ciampi, Till & Boyd, 1982; von Winterfeldt, 1975), most applications of multiattribute models or decision analysis assume that people know their own values and that the methods are unbiased channels for translating subjective feelings into analytically usable expressions. Fischhoff et al. argued that the strong effects of framing and information-processing considerations, acting upon inchoate preferences, can make elicitation procedures major forces in shaping the expression of values. In such cases, the method becomes the message. As shown above, subtle aspects of how problems are posed, questions are phrased, and responses are elicited can have a substantial effect on people's expressed preferences.

Managing Preferences

There are two potential reactions to the problems posed by labile values, one conservative and one radical. The conservative (decision theoretic) response assumes that true expressions of value are possible and attempts to clarify them through education (to reduce the uncertainty surrounding preferences) and the use of sophisticated elicitation techniques (to reduce biases). Consider, for example, a physician attempting to help a patient with cancer of the larynx choose between surgery and radiation therapy. Surgery produces longer life expectancy, but carries with it the loss of normal speech. Radiation therapy creates nausea and hair loss, but entails much lower risk of serious long-term side effects (for those who survive the cancer). The conservative approach attempts to assess utility functions for varying lengths of survival with and without normal speech—perhaps by asking the patient to assign certainty equivalents to gambles involving death and non-normal speech as outcomes. The patient's difficulty in forecasting how he or she would adapt to artificial speech or radiation therapy means that some education would have to take place prior to the value assessment procedure. That education might include contact with persons who did and did not choose surgery. How did these people react to the consequences of their decision? Did they correctly anticipate what it would be like to live without normal speech? Would they make the same decision again? After the education is completed, multiple assessment techniques would be employed to ensure that the patient's utility functions are faithfully captured.

The radical reaction to lability is to abandon the decision analytic approach on the grounds that it seeks to determine utility functions that do not exist and, as a result, has false pretensions about being able to identify the optimal decision. In the example of laryngeal cancer, decision analysis could produce utility functions that do not truly represent the patient's concerns, leading to recommendations of actions that are not in the patient's best interests. Furthermore, the very analytical process might raise the patient's anxiety about doing the right thing and increase the chances for strong post-decision regret. One possible alternative approach begins with the same educational effort, but then asks directly "which option do you prefer?" Patient and physician would then sift and weigh alternative reasons (or justifications), trying to develop a rationale for action. A strong rationale might buffer the patient from post-decision regret and make it easier to accept the consequences of the decision. If the patient is an intuitive decision theorist, this process could involve utility functions and maximization rules. However, quite different justifications could be equally legitimate if they have been thoughtfully derived.

Both education to inform preference and justification structuring to define it are forms of deliberate preference management. We manage our preferences in many ways. Aware, to some extent, of our multiple selves and changing tastes, we do such things as join Christmas clubs which bind us to our current preferences (Thaler & Shefrin, 1981), much as Ulysses forced his crew to tie him to the mast so that he might withstand the lure of the Sirens.

Deeper understanding of framing effects, which used-car salespeople have had for a long time and psychologists are beginning to acquire, could help us manage our own preferences more effectively (Thaler, 1983). Suppose, for example, that a person with $5,500 in a bank account misplaces a $100 bill. Rather than isolating and dwelling on this painful loss, assimilating it into one's total account may ease the sting by exploiting the perception that $5,500 is not that different from $5,600. Because neither perspective on the loss is inherently the "right" one, the choice between them could be a strategic decision, dependent upon the circumstances. If it is important to ensure that the mistake does not recur, then it might be best to isolate the loss, so as to maximize its impact. If the loss could not have been prevented, or if its impact has been traumatic, then one might well bury it so as to move on to other decisions.

The concept of preference management reflects the deep interplay between descriptive phenomena and normative principles. Experimental study of decision processes appears to be forging a new conception of preference, one that may require serious restructuring of normative theories and approaches toward improving decision making.

References

Abelson, R. P. Psychological status of the script concept. American Psychologist, 1981, 36, 715-729.

Abelson, R. P. Script processing in attitude formation and decision making. In J. S. Carroll and J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, N. J.: Erlbaum, 1976.

Allais, M. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole americaine. Econometrica, 1953, 21, 503-546.

Anderson, N. H. Functional measurement and psychophysical judgment. Psychological Review, 1970, 77, 153-170.

Anderson, N. H. Algebraic models in perception. In E. C. Carterette and M. P. Friedman (Eds.), Handbook of perception (Vol. 2). New York: Academic Press, 1974.

Anderson, N. H. Foundations of information integration theory. New York: Academic Press, 1981.

Anderson, N. H. & Shanteau, J. C. Information integration in risky decision making. Journal of Experimental Psychology, 1970, 84, 441-451.

Arrow, K. J. Social choice and individual values. New York: Wiley, 1951.

Arrow, K. J. Risk perception in psychology and economics. Economic Inquiry, 1982, 20, 1-9.

Ashton, R. H. Human information processing in accounting. Sarasota, Florida: American Accounting Association, 1982.

Bar-Hillel, M. On the subjective probability of compound events. Organizational Behavior and Human Performance, 1973, 9, 396-406.

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

Bar-Hillel, M. & Fischhoff, B. When do base rates affect predictions? Journal of Personality and Social Psychology, 1981, 41, 671-680.

Barron, F. H. & Person, H. B. Assessment of multiplicative utility functions via holistic judgments. Organizational Behavior and Human Performance, 1979, 24, 147-166.

Beach, L. R. & Mitchell, T. R. A contingency model for the selection of decision strategies. Academy of Management Review, 1978, 3, 439-449.

Beach, L. R., Townes, B. D., Campbell, F. L. & Keating, G. W. Developing and testing a decision aid for birth planning decisions. Organizational Behavior and Human Performance, 1976, 15, 99-116.

Becker, G. M., DeGroot, M. H. & Marschak, J. Probabilities of choices among very similar objects. Behavioral Science, 1963, 8, 306-311.

Bell, D. Regret in decision making under uncertainty. Operations Research, 1982, 30, 961-981.

Benbasat, I. & Taylor, R. N. Behavioral aspects of information processing for the design of management information systems. IEEE Transactions on Systems, Man and Cybernetics, 1982, SMC-12, 439-450.

Bernoulli, D. Specimen theoriae novae de mensura sortis. Commentarii Academiae Scientiarum Imperiales Petropolitanae, 1738, 5, 175-192. (Trans. by L. Sommer in Econometrica, 1954, 22, 23-36.)

Bettman, J. R. An information processing theory of consumer choice. Reading, Mass.: Addison-Wesley, 1979.

Bettman, J. R. & Jacoby, J. Patterns of processing in consumer information acquisition. In Beverlee B. Anderson (Ed.), Advances in consumer research. Chicago: Association for Consumer Research, 1976.

Beyth-Marom, R. The subjective probability of conjunctions. Decision Research Report 81-12, 1981.

Beyth-Marom, R. How probable is probable? Numerical translation of verbal probability expressions. Journal of Forecasting, 1982, 1, 257-269 (a).

Beyth-Marom, R. Perception of correlation reexamined. Memory and Cognition, 1982, 10(6), 511-519 (b).

Beyth-Marom, R. A curriculum on thinking under uncertainty. Instructional Science, in press.

Birnbaum, M. H. The devil rides again: Correlation as an index of fit. Psychological Bulletin, 1973, 79, 239-242.

Block, H. D. & Marschak, J. Random orderings and stochastic theories of responses. In I. Olkin et al. (Eds.), Contributions to probability and statistics. Stanford, Calif.: Stanford University Press, 1960.

Bouwman, M. J. The use of accounting information: Expert vs. novice behavior. In G. R. Ungson and D. N. Braunstein (Eds.), Decision making: An interdisciplinary inquiry. Boston: Kent, 1982.

Branscombe, A. Checkout on consumer math. American Education, 1975, 11 (Oct.), 21-24.

Brehmer, B. Preliminaries to a psychology of inference. Scandinavian Journal of Psychology, 1979, 20, 193-210.

Brehmer, B. Effect of cue validity on learning of complex rules in probabilistic inference tasks. Acta Psychologica, 1980, 44, 201-210 (a).

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

Brickman, P., Coates, D. & Janoff-Bulman, R. Lottery winners and accident victims: Is happiness relative? Journal of Personality and Social Psychology, 1978, 36, 917-927.

Bruner, J. S., Goodnow, J. J. & Austin, G. A. A study of thinking. New York: Wiley, 1956.

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

Brunswik, E. Perception and the representative design of experiments. Berkeley: University of California Press, 1956.

Bursztajn, H. & Hamm, R. M. The clinical utility of utility assessment. Medical Decision Making, 1982, 2, 161-165.

Cameron, P., Titus, D. G., Kostin, J. & Kostin, M. The life satisfaction of nonnormal persons. Journal of Counseling and Clinical Psychology, 1973, 41, 207-214.

Campbell, D. T. & Boruch, R. F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Bennett and A. A. Lumsdaine (Eds.), Evaluation and experiment: Some critical issues in assessing social programs. New York: Academic Press, 1975.

Capon, N. & Burke, M. Information seeking behavior in consumer durable purchase. In B. A. Greenberg and D. M. Bellenger (Eds.), Contemporary marketing thought. Chicago: American Marketing Association, 1977.

Chapman, L. J. & Chapman, J. P. Genesis of popular but erroneous psychodiagnostic observations. Journal of Abnormal Psychology, 1967, 73, 193-204.

Chew, S. H. & MacCrimmon, K. Alpha utility theory, lottery, composition, and the Allais Paradox. Working Paper #686, University of British Columbia, 1979.

Christensen-Szalanski, J. J. J. Problem-solving strategies: A selection mechanism, some implications, and some data. Organizational Behavior and Human Performance, 1978, 22, 307-323.

Christensen-Szalanski, J. J. J. Discount functions and the measurement of patients' values. Womens' decisions during childbirth. Medical Decision Making, in press.

Clarkson, G. P. E. Portfolio selection: A simulation of trust investment. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

Cohen, J. Can human irrationality be experimentally demonstrated? The Behavioral and Brain Sciences, 1981, 4, 317-331.

Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. Journal of Experimental Psychology, 1958, 55, 1-7.

Coombs, C. H. A theory of data. New York: Wiley, 1964.

Coombs, C. H. Portfolio theory: A theory of risky decision making. La Decision, Paris, France: Centre National de la Recherche Scientifique, 1969.

Coombs, C. H. Portfolio theory and the measurement of risk. In M. Kaplan and S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press, 1975.

Coombs, C. H. Risk preference and the theory of risk. In E. Lantermann and H. Feger (Eds.), _Similarity and choice_. Bern, Switzerland: Hans Huber, 1980, 346-364.

Coombs, C. H. & Bowen, J. Additivity of risk and portfolios. _Perception and Psychophysics_, 1971, _10_, 43-46.

Coombs, C. H., Dawes, R. M. & Tversky, A. _Mathematical psychology: An elementary introduction_. Englewood Cliffs, N. J.: Prentice-Hall, 1970.

Coombs, C. H. & Huang, L. C. Polynomial psychophysics of risk. _Journal of Mathematical Psychology_, 1970, _7_, 317-338 (a).

Coombs, C. H. & Huang, L. C. Tests of a portfolio theory of risk preference. _Journal of Experimental Psychology_, 1970, _85_, 23-29 (b).

Coombs, C. H. & Huang, L. C. Tests of the betweenness property of expected utility. _Journal of Mathematical Psychology_, 1976, _13_, 323-337.

Coombs, C. H. & Lehner, P. E. Evaluation of two alternative models for a theory of risk: I. Are moments of distributions useful in assessing risk? _Journal of Experimental Psychology: Human Perception and Performance_, 1981, _7_, 1110-1123.

Coombs, C. H. & Meyer, D. E. Risk-preference in coin-toss games. _Journal of Mathematical Psychology_, 1969, _6_, 514-527.

Coombs, C. H. & Pruitt, D. G. Components of risk in decision making: Probability and variance preferences. _Journal of Experimental Psychology_, 1960, _60_, 265-277.

Darlington, R. B. Multiple regression in psychological research and practice. _Psychological Bulletin_, 1968, _69_, 161-182.

Davidson, D., Suppes, P. & Siegel, S. _Decision making: An experimental approach_. Stanford, Calif.: Stanford University, 1957.

Dawes, R. M. A case study of graduate admissions: Applications of three principles of human decision making. American Psychologist, 1971, 26, 180-188.

Dawes, R. M. Shallow psychology. In J. S. Carroll and J. W. Payne (Eds.), Cognition and social behavior. Potomac, Md.: Erlbaum, 1976.

Dawes, R. M. Predictive models as a guide to preference. IEEE Transactions on Systems, Man and Cybernetics, 1977, SMC-7, 355-358.

Dawes, R. M. & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Debreu, G. Review of R. D. Luce. American Economic Review, 1960, 50, 186-188 (a).

Debreu, G. Topological methods in cardinal utility theory. In K. J. Arrow, S. Karlin and P. Suppes (Eds.), Mathematical methods in the social sciences, 1959. Stanford, Calif: Stanford University Press, 1960 (b).

de Finetti, B. Foresight: Its logical laws, its subjective sources. In H. E. Kyburg, Jr. and H. E. Smokler (Eds.), Studies in subjective probability. Huntington, New York: Krieger, 1980. (Originally published, 1937.)

de Finetti, B. Probability: Beware of falsifications. In H. E. Kyburg, Jr. and H. E. Smokler (Eds.), Studies in subjective probability. Huntington, New York: Krieger, 1980.

Diaconis, P. & Zabell, S. L. Updating subjective probability. Journal of the American Statistical Association, 1982, 77, 822-829.

Doherty, M. E., Mynatt, C. R., Tweney, R. D. & Schiavo, M. D. Pseudodiagnosticity. Acta Psychologica, 1979, 43, 111-121.

Dostoevsky, F. The gambler (A. R. MacAndrew, trans.). New York: Bantam Books, 1964. (Originally published, 1866.)

DuCharme, W. M. & Donnell, M. L. Intrasubject comparison of four response modes for "subjective probability" assessment. Organizational Behavior and Human Performance, 1973, 10, 108-117.

Dudyche, A. L. & Naylor, J. C. Characteristics of the human inference process in complex choice behavior situations. Organizational Behavior and Human Performance, 1966, 1, 110-128.

Eckenrode, R. T. Weighting multiple criteria, Management Science, 1965, 12, 180-192.

Edwards, W. Probability-preferences in gambling. American Journal of Psychology, 1953, 66, 349-364.

Edwards, W. Probability preferences among bets with differing expected values. American Journal of Psychology, 1954, 67, 56-67 (a).

Edwards, W. The reliability of probability preferences. American Journal of Psychology, 1954, 67, 68-95 (b).

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

Edwards, W. The prediction of decisions among bets. Journal of Experimental Psychology, 1955, 51, 201-214.

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

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

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

Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. Psychological Bulletin, 1970, 73, 211-230.

Einhorn, H. J. & Hogarth, R. M. Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 1978, 85, 395-416.

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

Ellsberg, D. Risk, ambiguity, and the Savage axioms. Quarterly Journal of Economics, 1961, 75, 643-669.

Ericsson, K. A. & Simon, H. A. Verbal reports as data. Psychological Review, 1980, 87, 215-251.

Estes, W. K. The cognitive side of probability learning. Psychological Review, 1976, 83, 37-64.

Evans, J. St. B. T. The psychology of deductive reasoning. London: Routledge & Kegan Paual, 1982.

Evans, J. St. B. T. & Pollard, P. Statistical judgment: A further test of the representativeness construct. Acta Psychologica, 1982, 51, 91-103.

Farquhar, P. H. Utility assessment methods. Working Paper 81-5, Graduate School of Administration, Davis, Calif.: University of California, 1982.

Feather, N. (Ed.) Expectations and actions: Expectancy-value models in psychology. Hillsdale, N.J.: Erlbaum, 1982.

Ferrell, W. R. & McGoey, P. J. A. A model of calibration for subjective probabilities. Organizational Behavior and Human Performance, 1980, 26, 32-53.

Fischer, G. W. Experimental applications of multi-attribute utility models. In D. Wendt and C. Vlek (Eds.), Utility, probability, and Human decision making. Dordrecht, The Netherlands: D. Reidel, 1975.

Fischer, G. W. Convergent validation of decomposed multiattribute utility assessment procedures for risky and riskless decisions. Organizational Behavior and Human Performance, 1977, 18, 295-315.

Fischhoff, B. Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 288-299.

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

Fischhoff, B. For those condemned to study the past. In D. Kahneman, P. Slovic and A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge, 1982 (b).

Fischhoff, B. Predicting frames. Journal of Experimental Psychology: Learning, Memory and Cognition, 1983, 9, 103-116.

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

Fischhoff, B., Goitein, B. & Shapira, Z. The experienced utility of expected utility approaches. In N. Feather (Ed.), Expectancy, incentive, and action. Hillsdale, N. J.: Erlbaum, 1982.

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L. & Keeney, R. L. Acceptable risk. New York: Cambridge, 1981.

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

Fischhoff, B., Slovic, P. & Lichtenstein, S. Fault trees: Sensitivity of estimated failure probabilities to problem representation. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 330-344.

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

Fishburn, P. C. The foundations of expected utility theory. Dordrecht, Holland: D. Reidel, 1982 (a).

Fishburn, P. C. Nontransitive measurable utility. Journal of Mathematical Psychology, 1982, 26, 31-67 (b).

Fishburn, P. C. Normative theories of decision making under risk and under uncertainty. Research paper, 75th Anniversary Colloquium Series, Harvard Business School, Cambridge, Massachusetts, 1983.

Fisher, I. The nature of capital and income. New York: Macmillan, 1906.

Flaherty, E. G. The thinking aloud technique and problem solving ability. Journal of Educational Research, 1975, 68, 223-225.

Friedman, M. P. Consumer confusion in the selection of supermarket products. Journal of Applied Psychology, 1966, 50, 529-534.

Furby, L. Interpreting regression toward the mean in developmental research. Developmental Psychology, 1973, 8, 172-179.

Llewellyn-Thomas, M., Sutherland, H. J., Tibshirani, R., Ciampi, A., Till, J. E. & Boyd, N. F. The measurement of patients' values in medicine. Medical Decision Making, 1982, 2, 449-462.

Loomes, G. & Sugden, R. Regret theory: An alternative theory of rational choice under uncertainty. The Economic Journal, 1982, 92, 805-824.

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

Lopes, L. L. Procedural debiasing. Wisconsin Human Information Processing Program, Report WHIPP 15, Dept. of Psychology, Madison, Wisconsin, 1982b.

Lopes, L. L. Toward a procedural theory of judgment. Wisconsin Human Information Processing Program, Report WHIPP 17, Dept. of Psychology, Madison, Wisconsin, 1982c.

Luce, R. D. Individual choice behavior. New York: Wiley, 1959.

Luce, R. D. The choice axiom after twenty years. Journal of Mathematical Psychology, 1977, 15, 215-233.

Luce, R. D. Several possible measures of risk. Theory and Decision, 1980, 12, 217-228.

Luce, R. D. Correction to "Several possible measures of risk." Theory and Decision, 1981, 13, 381.

Luce, R. D. & Suppes, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), Handbook of mathematical psychology, Vol. 3. New York: Wiley, 1965.

Luce, R. E. & Tukey, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology, 1964, 1, 1-27.

MacCrimmon, K. R. Descriptive and normative implications of the decision-theory postulates. In K. Borch and J. Mossin (Eds.), Risk and uncertainty. New York: St. Martins Press, 1968.

MacCrimmon, K. R. & Larsson, S. Utility theory: Axioms versus "paradoxes." In M. Allais and O. Hagen (Eds.), Rational decisions under uncertainty, special volume of Theory and Decision, 1976.

MacCrimmon, K. R. & Siu, J. K. Making trade-offs. Decision Sciences, 1974, 5, 680-704.

March, J. G. Bounded rationality, ambiguity, and the engineering of choice. The Bell Journal of Economics, 1978, 9, 587-608.

March, J. G. Theories of choice and the making of decisions. Presented at the American Association for the Advancement of Science, Washington, D. C., January 5, 1982.

Matheson, J. E. & Roths, W. J. Decision analysis of space projects: Voyager Mars. In R. Howard, J. E. Matheson and K. L. Miller (Eds.), Readings in decision analysis. Palo Alto, Calif.: Stanford Research Institute, 1967.

McNeil, B. J., Pauker, S. G., Sox, H. C. Jr. & Tversky, A. On the elicitation of preferences for alternative therapies. The New England Journal of Medicine, 1982, 306, 1259-1262.

McNeil, B. J., Weichselbaum, R. & Pauker, S. G. Speech and survival: Tradeoffs between quality and quantity of life in laryngeal cancer. The New England Journal of Medicine, 1981, 305, 982-987.

Meehl, P. E. & Rosen, A. Antecedent probability and the efficacy of psychometric signs, patterns or cutting scores. Psychological Bulletin, 1955, 52, 194-216.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Montgomery, H. Decision rules and the search for a dominance structure: Towards a process model of decision making. In P. Humphreys, O. Svenson and A. Vari (Eds.), Analyzing and aiding decision processes. Amsterdam: North Holland, 1983.

Moskowitz, H. Effects of problem representation and feedback on rational behavior in Allais and Morlat-type problems. Decision Sciences, 1974, 3, 225-241.

Mosteller, F. & Nogee, P. An experimental measurement of utility. Journal of Political Economics, 1951, 59, 371-404.

Munera, H. A. & de Neufville, R. A decision analysis model when the substitution principle is not acceptable. Paper presented to the First International Conference on Foundations of Utility and Risk Theory. Oslo, Norway, July, 1982.

Murphy, A. H. & Winkler, R. C. Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? National Weather Digest, 1977, 2(2), 2-9.

National Academy of Sciences. Research need in human factors. Washington, D.C.: The Academy, 1983.

Newell, A. Reasoning, problem solving and decision processes: The problem space as a fundamental category. In R. S. Nickerson (Ed.), Attention and Performance, VIII. Hillsdale, N. J.: Erlbaum, 1980.

Newell, A. & Simon, H. A. Human problem solving. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

Nisbett, A. & Ross, L. Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, N.J.: Prentice-Hall, 1980.

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

Novick, M. R., Dekeyrel, D. F. & Chuang, D. T. Local and regional coherence utility procedures, 1981.

Pauker, S. G. Coronary artery surgery: The use of decision analysis. Annals of Internal Medicine, 1976, 85, 8-18.

Payne, J. W. Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance, 1976, 16, 366-387.

Payne, J. W. Information processing theory: Some concepts applied to decision research. In T. Wallsten (Ed.), Cognitive processes in choice and decision behavior. Hillsdale, N.J.: Erlbaum, 1980.

Payne, J. W. Contingent decision behavior. Psychological Bulletin, 1982, 92, 382-401.

Payne, J. W. & Braunstein, M. L. Preferences among gambles with equal underlying distributions. Journal of Experimental Psychology, 1971, 87, 13-18.

Payne, J. W. & Braunstein, M. L. Risky choice: An examination of information acquisition behavior. Memory and Cognition, 1978, 6, 554-561.

Payne, J. W., Braunstein, M. L. & Carroll, J. S. Exploring predecisional behavior: An alternative approach to decision research. Organizational Behavior and Human Performance, 1978, 22, 17-44.

Payne, J. W., Laughhunn, D. J. & Crum, R. Translation of gambles and aspiration level effects in risky choice behavior. Management Science, 1980, 26, 1039-1060.

Payne, J. W., Laughhunn, D. J. & Crum, R. Further tests of aspiration level effects in risky choice behavior. Management Science, 1981, 27, 953-958.

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

Pitz, G. F. Subjective probability distributions for imperfectly known quantities. In L. W. Gregg (Ed.), Knowledge and Cognition. Potomac, Maryland: Erlbaum, 1974.

Pitz, G. F. Decision making and cognition. In H. Jungermann and G. de Zeeuw (Eds.), Decision making and change in human affairs. Dordrecht, Holland: D. Reidel, 1977.

Pitz, G. F., Sachs, N. J. & Heerboth, J. Procedures for eliciting choices in the analysis of individual decisions. Organizational Behavior and Human Performance, 1980, 26, 396-408.

Pliske, R. M., Gettys, C. F., Manning, L. & Casey, J. T. Act generation performance: The effects of incentive. TR 15-8-82. Decision Processes Laboratory, University of Oklahoma, Norman, Oklahoma, 1982.

Polanyi, M. Personal knowledge. London: Routledge & Kegan Paul, 1958.

Posner, M. I. Protocol analysis and human cognition. In G. Ungson and D. Braunstein (Eds.), Decision making: An interdisciplinary inquiry. Boston: Kent, 1982.

Poulton, E. C. Biases in quantitative judgments. Applied Ergonomics, 1982, 13, 31-42.

Raaij, W. F. van. Techniques for process training in decision making. In L. Sjöberg, T. Tyszka and J. Wise (Eds.), Human decision making. Bodafors, Sweden: Doxa, 1983.

Raiffa, H. Decision analysis. Reading, Mass.: Addison-Wesley, 1968.

Ramsey, F. P. Truth and probability. In F. P. Ramsey, The foundations of mathematics and other logical essays. New York: Harcourt Brace, 1931.

Restle, R. Psychology of judgment and choice: A theoretical essay. New York: Wiley, 1961.

Robinson, G. H. Continuous estimation of a time-varying probability. Ergonomics, 1964, 7, 7-21.

Rumelhart, D. L. & Greeno, J. G. Similarity between stimuli: An experimental test of the Luce and Restle choice models. Journal of Mathematical Psychology, 1971, 8, 370-381.

Russ, F. A. Consumer evaluation of alternative product models. Unpublished doctoral dissertation, Carnegie-Mellon University, 1971.

Russo, J. E. The value of unit-price information. Journal of Marketing Research, 1977, 14, 193-201.

Russo, J. E. & Dosher, B. A. Dimensional evaluation: A heuristic for binary choice. Unpublished working paper. San Diego: Department of Psychology, University of California, 1975.

Russo, J. E. & Rosen, L. D. An eye-fixation analysis of multialternative choice. Memory and Cognition, 1975, 3, 267-276.

Sage, A. P. & White, E. B. Decision and information structures in regret models of judgment and choice. IEEE Transactions on Systems, Man and Cybernetics, 1983, SMC-13, 136-145.

Samuelson, P. A. Probability and the attempts to measure utility. The Economic Review (Toyko, Japan), 1950, 1, 167-173.

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

Schelling, T. C. Identifying that authentic self. Paper presented at the meeting of the working group on rationality. Paris, Janauary 4-6, 1982.

Schoemaker, P. J. H. & Waid, C. C. An experimental comparison of different approaches to determining weights in additive utility models. Management Science, 1982, 28, 182-196.

Schum, D. Current developments in research on cascaded inference processes. In T. Wallsten (Ed.), Cognitive processes in choice and decision behavior. Hillsdale, N.J.: Erlbaum, 1980.

Shafer, G. & Tversky, A. Weighing evidence: The design and comparison of probability thought experiments. Theory and Decision, in press.

Shaklee, H. & Tucker, D. A rule analysis of judgments of covariation between events. Memory and Cognition, 1980, 8, 459-467.

Shapira, Z. Measuring subjective probabilities by the magnitude production method. Organizational Behavior and Human Performance, 1975, 14, 314-320.

Sheridan, J. E., Richards, M. D. & Slocum, J. W. Comparative analysis of expectancy and heuristic models of decision behavior. Journal of Applied Psychology, 1975, 60, 361-368.

Shuford, E. H. Jr. Percentage estimation of proportion as a function of element type, exposure type, and task. Journal of Experimental Psychology, 1961, 61, 430-436.

Shuford, E. H. Jr., Albert, A. & Massengill, H. E. Admissible probability measurement procedures. Psychometrika, 1966, 31, 125-145.

Shugan, S. M. The cost of thinking. Journal of Consumer Research, 1980, 7, 99-111.

Simon, H. A. Rational choice and the structure of the environment. Psychological Review, 1956, 63, 129-138.

Simon, H. A. Theories of decision making in economics and behavioral science. American Economic Review, 1959, 49, 253-283.

Simon, H. A. Discussion: Cognition and social behavior. In J. S. Carroll and J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, N. J.: Erlbaum, 1976.

Sjöberg, L. Choice frequency and similarity. Scandinavian Journal of Psychology, 1977, 18, 103-115.

Slovic, P. From Shakespeare to Simon: Speculation--and some evidence--about man's ability to process information. ORI Research Monograph, 1972, 12(2).

Slovic, P. Choice between equally valued alternatives. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 280-287.

Slovic, P. & Fischhoff, B. On the psychology of experimental surprises. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 544-551.

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

Slovic, P., Fischhoff, B. & Lichtenstein, S. Response mode, framing, and information processing effects in risk assessment. In R. M. Hogarth (Ed.), New directions for methodology of social and behavioral science: The framing of questions and the consistency of response. San Francisco: Jossey-Bass, 1982.

Slovic, P. & Lichtenstein, S. The importance of variance preferences in gambling decisions. Journal of Experimental Psychology, 1968, 78, 646-654 (a).

Slovic, P. & Lichtenstein, S. The relative importance of probabilities and payoffs in risk-taking. Journal of Experimental Psychology Monograph Supplement, 1968, 78(3), part 2 (b).

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

Slovic, P. & Lichtenstein, S. Preference reversals: A broader perspective. American Economic Review, 1983, 73, 596-605.

Slovic, P., Lichtenstein, S. & Edwards, W. Boredom-induced changes in preferences among bets. American Journal of Psychology, 1965, 78, 208-217.

Slovic, P. & MacPhillamy, D. J. Dimensional commensurability and cue utilization in comparative judgment. Organizational Behavior and Human Performance, 1974, 11, 172-194.

Slovic, P. & Tversky, A. Who accepts Savage's axiom? Behavioral Science, 1974, 19, 368-373.

Smedslund, J. The concept of correlation in adults. Scandinavian Journal of Psychology, 1963, 4, 165-173.

Snyder, M. & Swann, W. B. Hypothesis testing procedures in social interaction. Journal of Personality and Social Psychology, 1978, 36, 1202-1212.

Spetzler, C. S. & Staël von Holstein, C. A. S. Probability encoding in decision analysis. Management Science, 1975, 22, 340-358.

Stillwell, W. G., Barron, F. H. & Edwards, W. Evaluating credit applications: A valuation of multiattribute utility techniques against a real world criterion. Organizational Behavior and Human Performance, 1983, 32, 87-108.

Stillwell, W. G., Seaver, D. A. & Edwards, W. A comparison of weight approximation techniques in multiattribute utility decision making. Organizational Behavior and Human Performance, 1981, 28, 62-77.

Svenson, O. Coded think-aloud protocols obtained when making a choice to purchase one of seven hypothetically offered houses. Unpublished manuscript, University of Stockholm, 1974.

Svenson, O. Process descriptions of decision making. Organizational Behavior and Human Performance, 1979, 23, 86-112.

Thaler, R. Using mental accounting in a theory of consumer behavior. Graduate School of Business and Public Administration, Working Paper 83-02, revised May 1983.

Thaler, R. H. & Shefrin, H. M. An economic theory of self control. Journal of Political Economy, 1981, 89, 392-406.

Thurstone, L. L. A law of comparative judgment. Psychological Review, 1927, 34, 273-286.

Tichener, E. B. A textbook of psychology. New York: MacMillan, 1910.

Tucker, L. R. A suggested alternative formulation in the development of Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. Psychological Review, 1964, 71, 528-530.

Tversky, A. Additivity, utility and subjective probability. Journal of Mathematical Psychology, 1967, 4, 175-202 (a).

Tversky, A. Utility theory and additivity analysis of risky choices. Journal of Experimental Psychology, 1967, 75, 27-36 (b).

Tversky, A. Intransitivity of preferences. Psychological Review, 1969, 76, 31-48.

Tversky, A. Choice by elimination. Journal of Mathematical Psychology, 1972, 9, 341-367 (a).

Tversky, A. Elimination by aspects: A theory of choice. Psychological Review, 1972, 79, 281-299 (b).

Tversky, A. & Kahneman, D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 1973, 4, 207-232.

Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Tversky, A. & Kahneman, D. The framing of decisions and the psychology of choice. Science, 1981, 211, 453-458.

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

Tversky, A. & Kahneman, D. Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, in press.

Tversky, A. & Russo, J. E. Similarity and substitutability in binary choices. Journal of Mathematical Psychology, 1969, 6, 1-12.

Tversky, A. & Sattath, S. Preference trees. Psychological Review, 1979, 86, 542-573.

Tyszka, T. Simple decision strategies vs. multi-attribute utility theory approach to complex decision problems. Praxiology Yearbook #2, 1981.

Tyszka, T. Contextual multiattribute decision rules. In L. Sjöberg, T. Tyszka and J. Wise (Eds.), Human decision making. Bodafors, Sweden: Doxa, 1983.

von Neumann, J. & Morgenstern, O. Theory of games and economic behavior. Princeton, N. J.: Princeton University Press, 1947.

von Winterfeldt, D. An overview, integration, and evaluation of utility theory for decision analysis. SSRI Technical Research Report 75-9. Los Angeles: University of Southern California, 1975.

Wagenaar, W. A. Generation of random sequences by human subjects: A critical survey of literature. Psychological Bulletin, 1972, 77, 65-72.

Wainer, H. Estimating coefficients in linear models: It don't make no never mind. Psychological Bulletin, 1976, 83, 213-217.

Wallsten, T. & Budescu, D. Encoding subjective probabilities: A psychological and psychometric review. Management Science, 1983, 29, 151-173.

Ward, W. & Jenkins, H. The display of information and the judgment of contingency. Canadian Journal of Psychology, 1965, 19, 231-241.

Wason, P. C. On the failure to eliminate hypotheses in a perceptual task. Quarterly Journal of Experimental Psychology, 1960, 12, 129-140.

Wason, P. C. Reasoning about a rule. Quarterly Journal of Experimental Psychology, 1968, 23, 273-281.

Weber, E. U. Combine and conquer: A joint application of conjoint and functional approaches to the problem of risk measurement. Journal of Experimental Psychology: Human Perception and Performance, in press.

White, G. F. Choice of adjustment to floods. Research Paper No. 93. Chicago: University of Chicago, Dept. of Geography, 1964.

Whitlow, J. W. Jr. & Estes, W. K. Judgments of relative frequency in relation to shifts of event frequencies: Evidence for a limited-capacity model. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 395-408.

Wright, P. L. The harassed decision maker: Time pressures, distraction, and the use of evidence. Journal of Applied Psychology, 1974, 59, 555-561(a).

Wright, P. L. The use of phased, noncompensatory strategies in decisions between multi-attribute products. Research Paper 223, Graduate School of Business, Stanford University, 1974 (b).

Wright, P. L. & Barbour, F. Phased decision strategies: Sequels to an initial screening. In M. K. Starr and M. Zeleny (Eds.), North-Holland/TIMS studies in the management sciences: Multiple criteria decision making. Amsterdam: North Holland, 1977.

Footnotes

1. The compelling nature of transitivity as a rational principle is illustrated by the fact that those who violate transitivity are behaving in effect as "money pumps." Suppose an individual prefers y to x, z to y, and x to z. It is reasonable to assume that he or she will pay some amount of money to replace x by y, a second amount to replace y by z and a third amount to replace z by x. Thus the individual ends up with the original alternative and less money.

2. The property of single-peakedness means here that, for any three gambles with the same EU and ordered in risk, the intermediate one cannot be the least preferred. Psychologically, this implies that, for a fixed expected value, an individual has an optimum level of risk and that preference declines as risk increases or decreases from this optimum.

3. The reader is referred to Chapter __ by Luce and Krumhansl who discuss these issues from a measurement perspective.

4. This distinction is not often made. The acronym MAUT (for multi-attribute utility theory) is often used to describe the models discussed here. Unfortunately, it is also sometimes used to decribe the risky multi-attribute models covered in the next section.

5. The form of this model is similar to prospect theory (Kahneman & Tversky, 1979), although the underlying rational and primitives are different.

Table 1 *Debiasing methods*

Assumption	Strategies
Faulty tasks	
Unfair tasks	Raise stakes
	Clarify instructions/stimuli
	Discourage second-guessing
	Use better response modes
	Ask fewer questions
Misunderstood tasks	Demonstrate alternative goal
	Demonstrate semantic disagreement
	Demonstrate impossibility of task
	Demonstrate overlooked distinction
Faulty judges	
Perfectible individuals	Warn of problem
	Describe problem
	Provide personalized feedback
	Train extensively
Incorrigible individuals	Replace them
	Recalibrate their responses
	Plan on error
Mismatch between judges and task	
Restructuring	Make knowledge explicit
	Search for discrepant information
	Decompose problem
	Consider alternative situations
	Offer alternative formulations
Education	Rely on substantive experts
	Educate from childhood

Source: Fischhoff, 1982a

Table 2

Matrix Form of a Simple Decision Problem

			State of nature	
			sun (E_1)	rain (E_2)
Alternatives	A_1	carry umbrella	stay dry carrying umbrella	stay dry carrying umbrella
	A_2	leave umbrella	dry and unburdened	wet and unburdened

Table 3
Gambles Used as Stimuli
by Tversky (1969)

Gamble	*Probability of Winning*	*Payoff (in \$)*	*Expected Value*
a	7/24	5.00	1.46
b	8/24	4.75	1.58
c	9/24	4.50	1.69
d	10/24	4.25	1.77
e	11/24	4.00	1.83

Table 4

An Exemplar Indifference Task:
Choose a Desert for Dinner B that Makes
You Indifferent Between Dinners A and B

	Dinner A	Dinner B
Soup	Chicken Gumbo	Cream of Asparagus
Salad	Tossed Greens	Spinach
Entree	Steak	Lobster
Vegetable	Baked Potato	Rice Pilaf
Desert	Chocolate Cake	?

Table 5. Protocol for a Subject Selecting among 12 Apartments

(A) Protocol	(A) Protocol (continued)
Let's just see what the rents are in all the apartments first.	Kitchen facilities in *A* are poor.
The rent of *A* is $140.	In *A*, poor.
The rent of *B* is $110.	In *B*, poor.
The rent of *C* is $170.	In *J*, fair.
...	In *H*, they're good.
Um, $170 is too much.	Oh, *J* and *H* have better kitchen facilities than *A* and *B*.
But, if the other ones aren't good, I'll look at them later.	And everything else is about the same.
But, right now I'll look at the other ones.	So eliminate those two.
...	And, decide between these two.
I'm going to look at landlord attitude.	...
In *H*, it's fair.	Let's see furniture quality.
In *D*, it's poor.	In *H*, it is below average.
B, it's fair, and	In *J*, it's below average, so that's about the same there.
A, it's good.	...
So, one of them . . . is poor.	Landlord attitude in *J* is better than in *H*.
So that's important to me.	...
...	In *J*, the rooms are larger,
So, I'm not going to live any place where where it's poor.	so, I guess, *J* will be better.
...	

Source: Payne, 1976

Figure Captions

1. A possible fault tree for discovering why a car won't start. Source: Fischhoff, Slovic & Lichtenstein, 1978.

2. Calibration curves showing overconfidence. Source: Lichtenstein, Fischhoff & Phillips, 1982.

3. The lens model. Source: Hammond et al., 1975.

4. Tree representation of a simple decision problem. Circles represent uncertain events; the square represents the decision point.

5. Decision tree representing the problem faced by a mining company trying to determine what to bid for two parcels of land with extensive ore deposits. Circles represent uncertain events; squares represent decision alternatives. Source: Hax & Wiig, 1977.

6. Matrix representation of the Allais problem.

7. Duplex and standard gambles used to study moment effects. Source: Slovic & Lichtenstein, 1968(a); Payne & Braunstein, 1971.

8. A hypothetical weighting function. Source: Kahneman & Tversky, 1979.

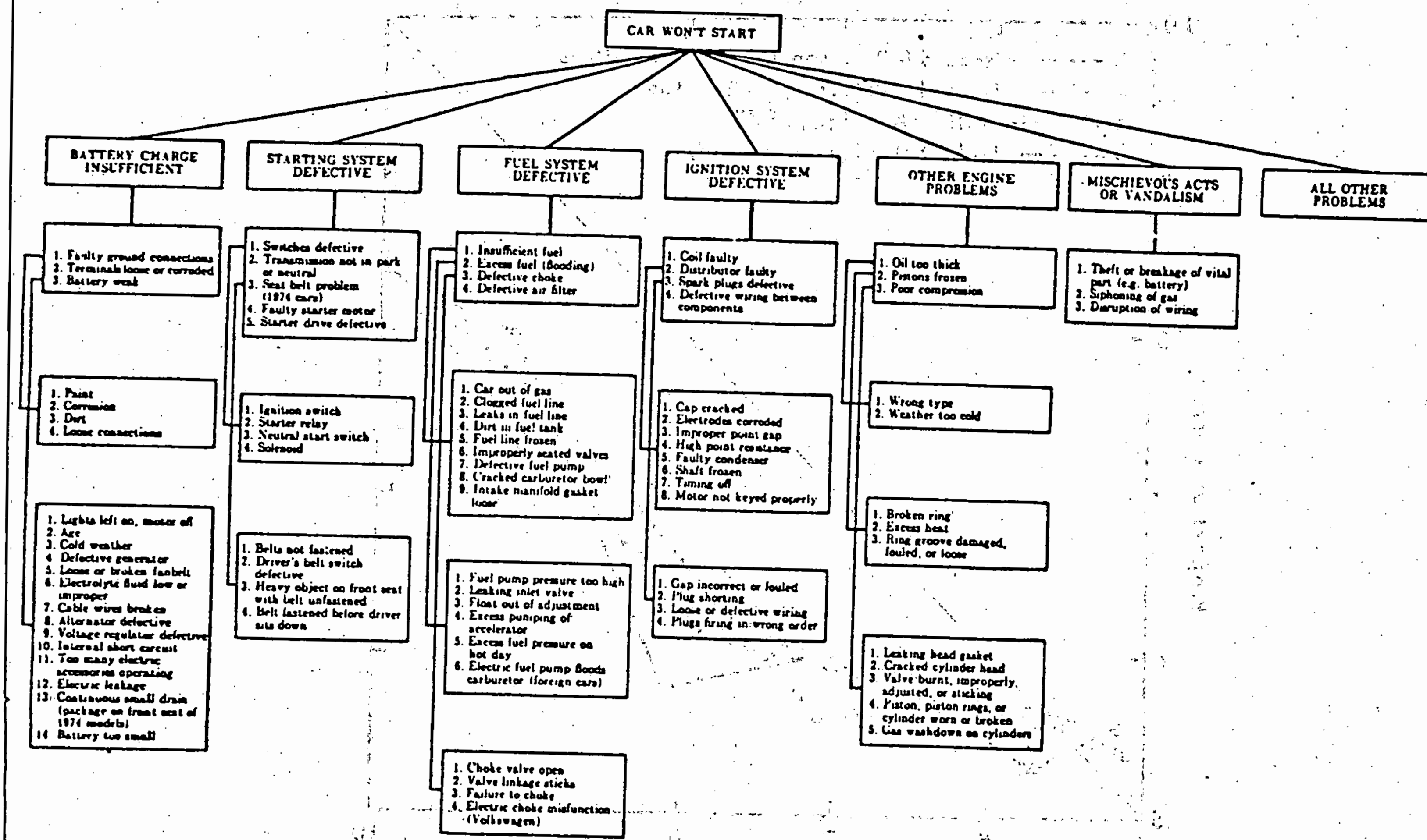
9. S-shaped value function. Source: Kahneman & Tversky, 1979.

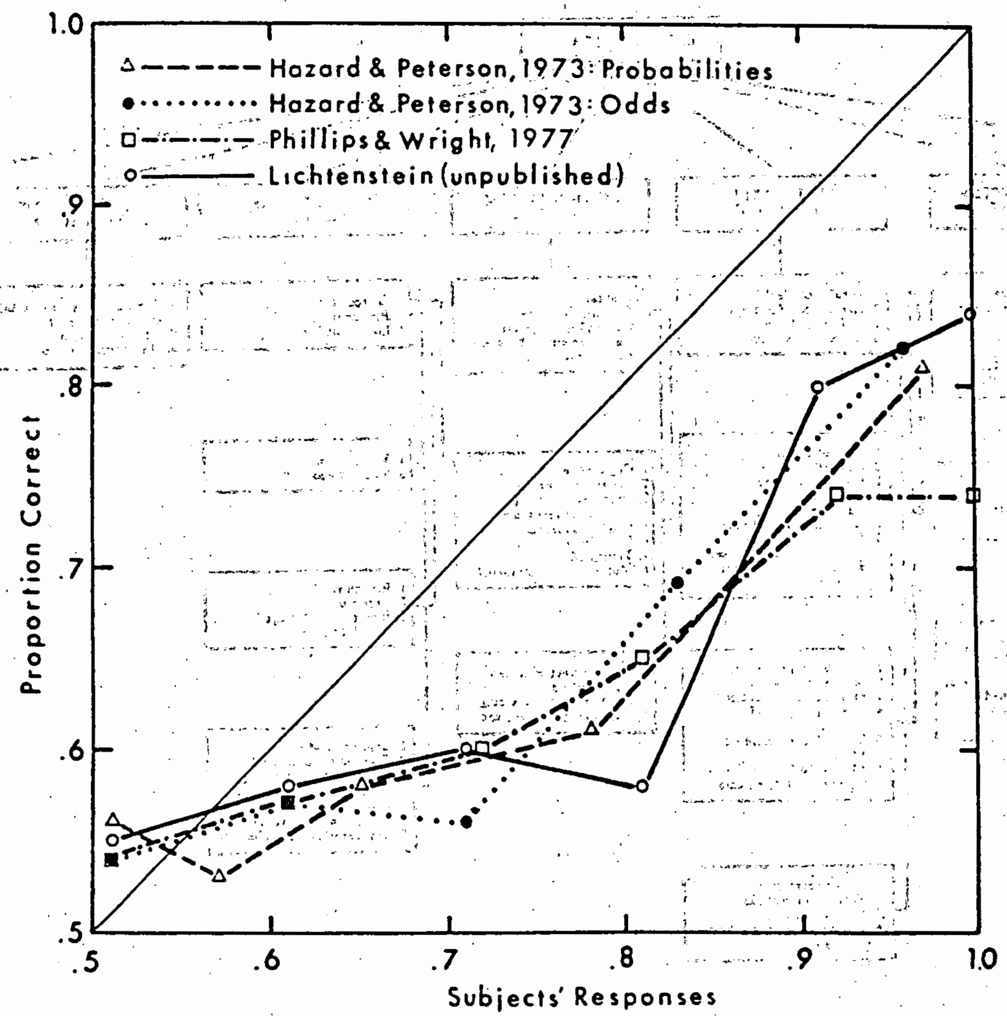
10. Value sets in choice between A and B. Source: Restle, 1961.

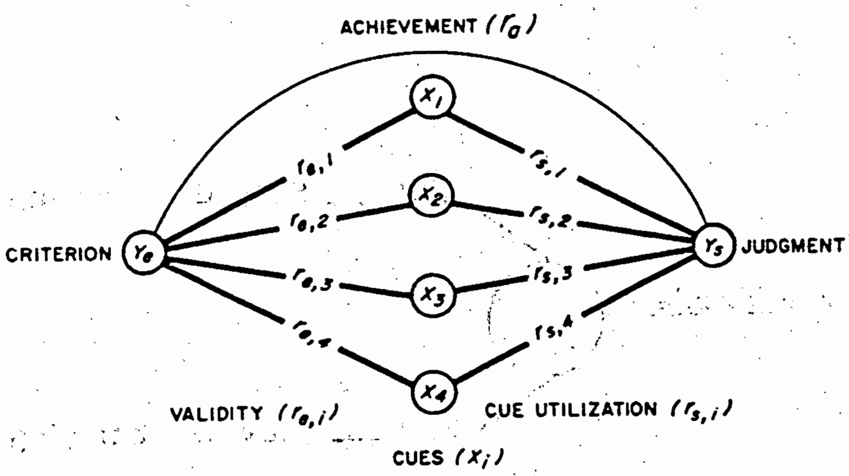
11. Tree representation of the choice among entrees. Source: Tversky & Sattath, 1979.

12. A weak-order indifference map. Objects A, B, C, and D are equal in utility. E is better than all of them, and F is worse.

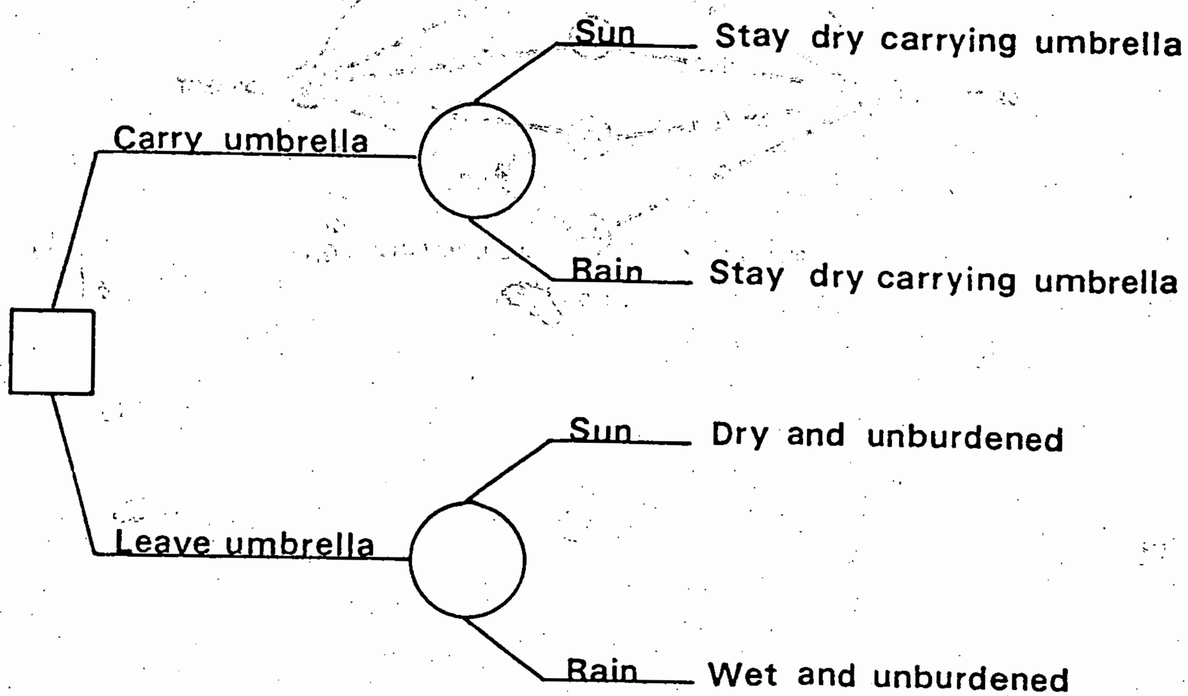
13. Clarkson's yield portfolio stock selection model. Source: Clarkson, 1962.

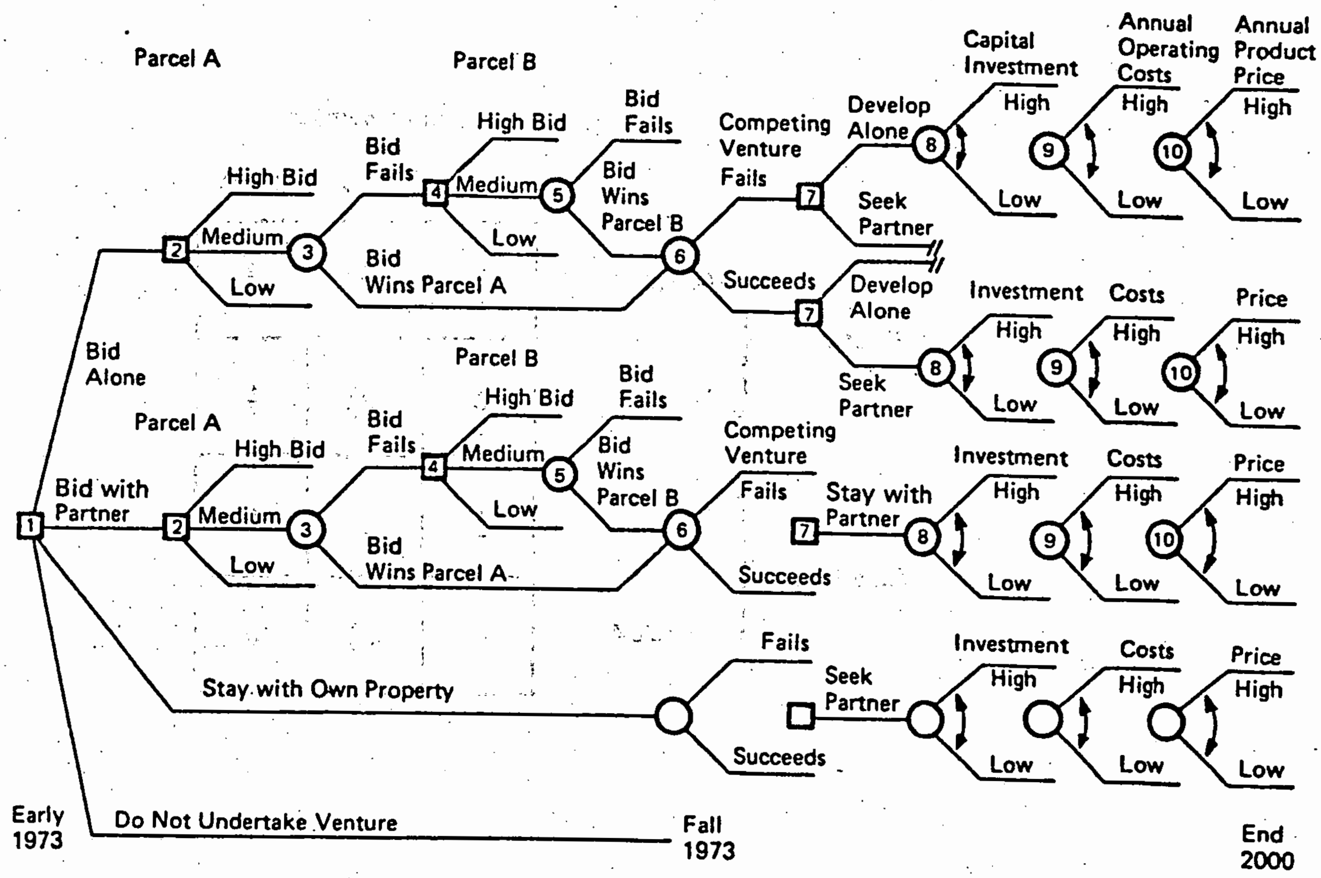

CAR WON'T START
BATTERY CHARGE INSUFFICIENT
1. Faulty ground connections
2. Terminals loose or corroded
3. Battery weak
1. Paint
2. Corrosion
3. Dirt
4. Loose connections
1. Lights left on, motor off
2. Age
3. Cold weather
4. Defective generator
5. Loose or broken fanbelt
6. Electrolyte fluid low or improper
7. Cable wires broken
8. Alternator defective
9. Voltage regulator defective
10. Internal short circuit
11. Too many electric accessories operating
12. Electric leakage
13. Continuous small drain (package on front seat of 1974 models)
14. Battery too small
STARTING SYSTEM DEFECTIVE
1. Switches defective
2. Transmission not in park or neutral
3. Seat belt problem (1974 cars)
4. Faulty starter motor
5. Starter drive defective
1. Ignition switch
2. Starter relay
3. Neutral start switch
4. Solenoid
1. Belts not fastened
2. Driver's belt switch defective
3. Heavy object on front seat with belt unfastened
4. Belt fastened before driver sits down
FUEL SYSTEM DEFECTIVE
1. Insufficient fuel
2. Excess fuel (flooding)
3. Defective choke
4. Defective air filter
1. Car out of gas
2. Clogged fuel line
3. Leaks in fuel line
4. Dirt in fuel tank
5. Fuel line frozen
6. Improperly seated valves
7. Defective fuel pump
8. Cracked carburetor bowl
9. Intake manifold gasket loose
1. Fuel pump pressure too high
2. Leaking inlet valve
3. Float out of adjustment
4. Excess pumping of accelerator
5. Excess fuel pressure on hot day
6. Electric fuel pump floods carburetor (foreign cars)
1. Choke valve open
2. Valve linkage sticks
3. Failure to choke
4. Electric choke misfunction (Volkswagen)
IGNITION SYSTEM DEFECTIVE
1. Coil faulty
2. Distributor faulty
3. Spark plugs defective
4. Defective wiring between components
1. Cap cracked
2. Electrodes corroded
3. Improper point gap
4. High point resistance
5. Faulty condenser
6. Shaft frozen
7. Timing off
8. Motor not keyed properly
1. Gap incorrect or fouled
2. Plug shorting
3. Loose or defective wiring
4. Plugs firing in wrong order
OTHER ENGINE PROBLEMS
1. Oil too thick
2. Pistons frozen
3. Poor compression
1. Wrong type
2. Weather too cold
1. Broken ring
2. Excess heat
3. Ring groove damaged, fouled, or loose
1. Leaking head gasket
2. Cracked cylinder head
3. Valve burnt, improperly adjusted, or sticking
4. Piston, piston rings, or cylinder worn or broken
5. Gas washdown on cylinders
MISCHIEVOUS ACTS OR VANDALISM
1. Theft or breakage of vital part (e.g. battery)
2. Siphoning of gas
3. Disruption of wiring
ALL OTHER PROBLEMS


Hazard & Peterson, 1973: Probabilities
Hazard & Peterson, 1973: Odds
Phillips & Wright, 1977
Lichtenstein (unpublished)
Proportion Correct
Subjects' Responses
1.0
.9
.8
.7
.6
.5
.5
.6
.7
.8
.9
1.0



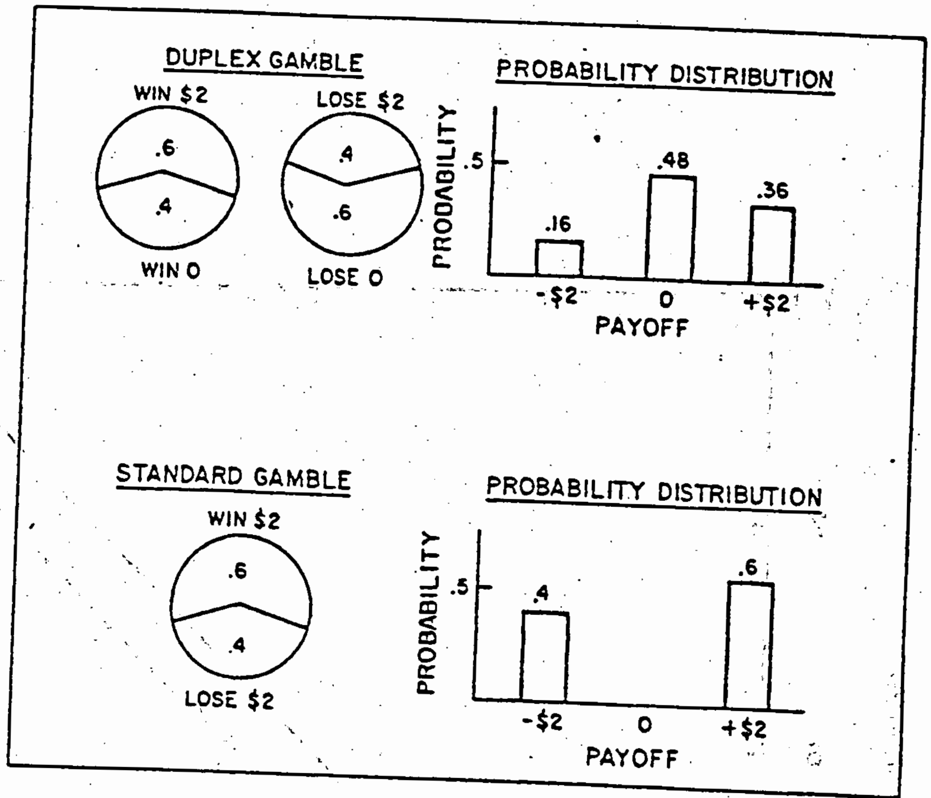
ACHIEVEMENT (r_a)
CRITERION
Y_e
X_1
X_2
X_3
X_4
$r_{e,1}$
$r_{e,2}$
$r_{e,3}$
$r_{e,4}$
$r_{s,1}$
$r_{s,2}$
$r_{s,3}$
$r_{s,4}$
Y_s
JUDGMENT
VALIDITY ($r_{e,i}$)
CUE UTILIZATION ($r_{s,i}$)
CUES (X_i)

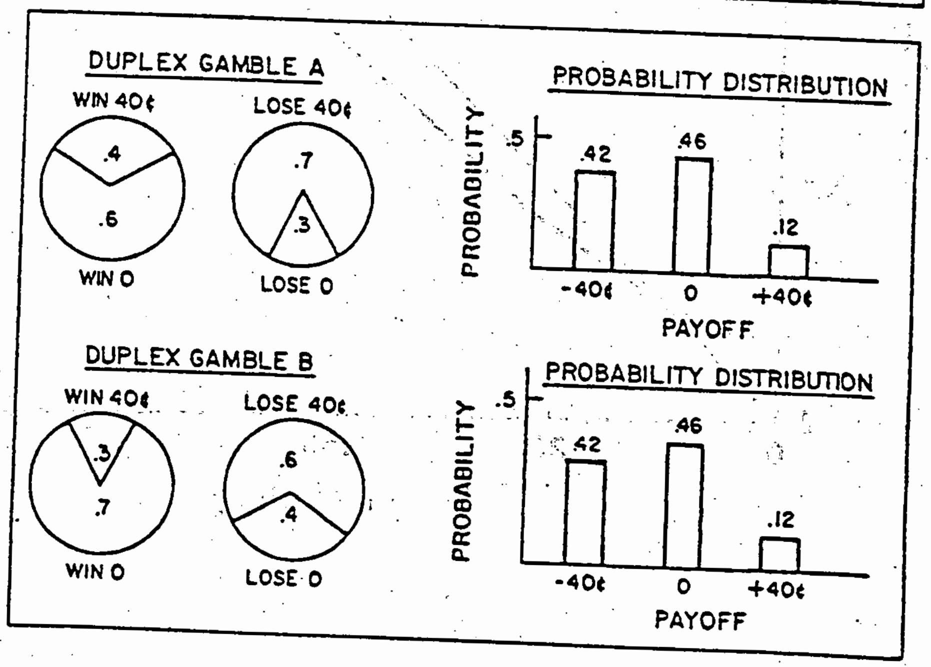

Carry umbrella
Sun
Stay dry carrying umbrella
Rain
Stay dry carrying umbrella
Leave umbrella
Sun
Dry and unburdened
Rain
Wet and unburdened

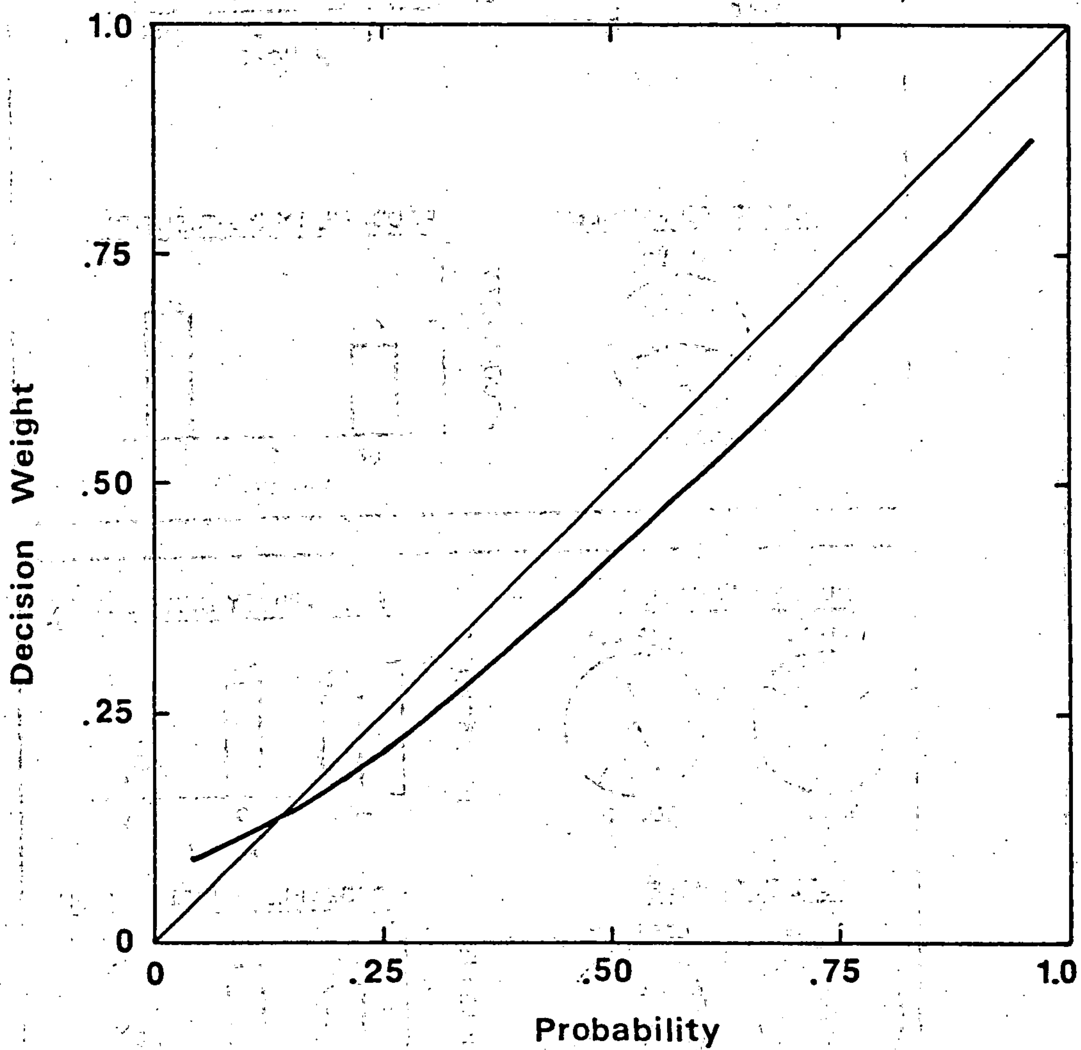


Parcel A
Parcel B
High Bid
Medium
Low
Bid Fails
Bid Wins Parcel A
High Bid
Medium
Low
Bid Fails
Bid Wins Parcel B
Competing Venture Fails
Succeeds
Develop Alone
Seek Partner
Capital Investment
High
Low
Annual Operating Costs
High
Low
Annual Product Price
High
Low
Investment
Costs
Price
Bid Alone
Bid with Partner
Stay with Partner
Stay with Own Property
Fails
Succeeds
Early 1973
Do Not Undertake Venture
Fall 1973
End 2000

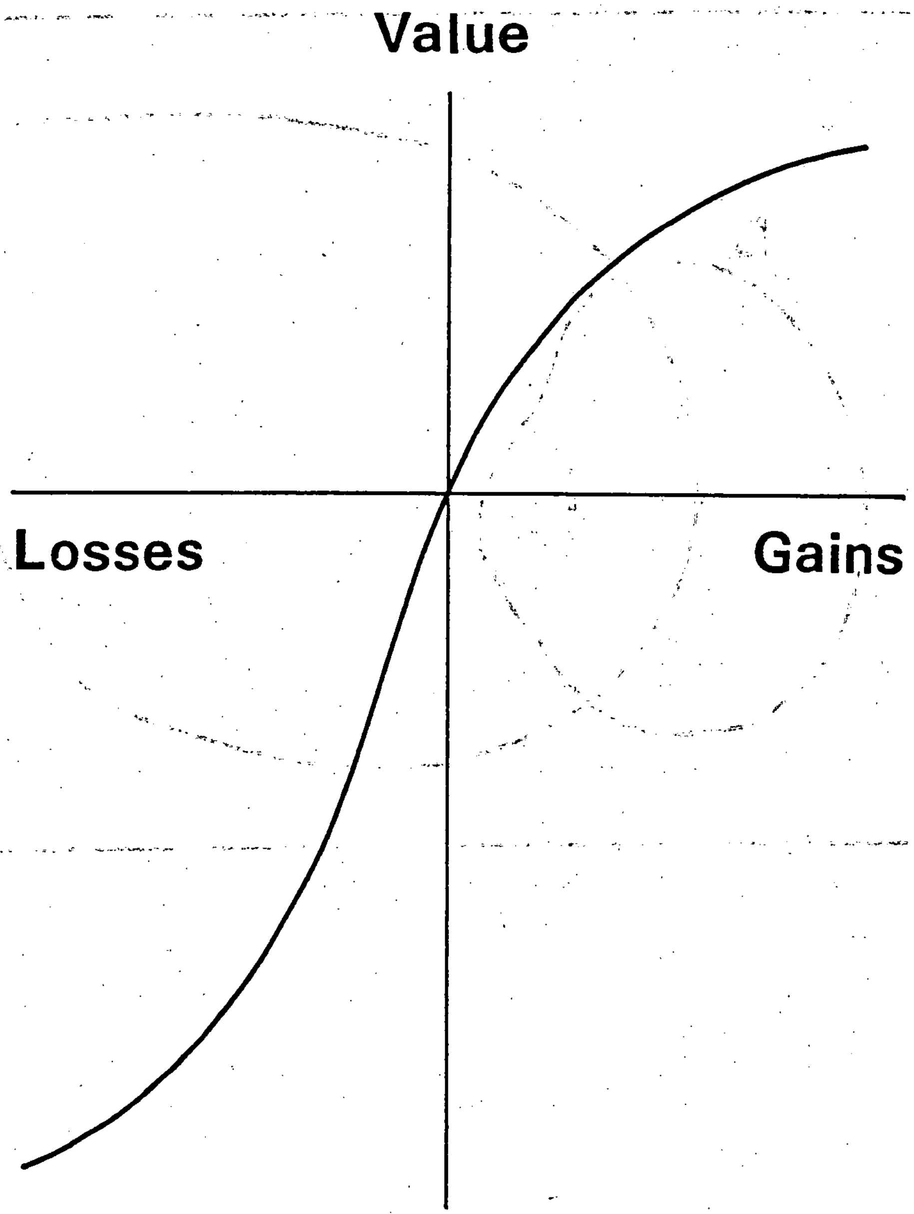
		Ticket number		
		1	2-11	12-100
Situation 1	Gamble 1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
	Gamble 2	0	$2\frac{1}{2}$	$\frac{1}{2}$
Situation 2	Gamble 3	$\frac{1}{2}$	$\frac{1}{2}$	0
	Gamble 4	0	$2\frac{1}{2}$	0

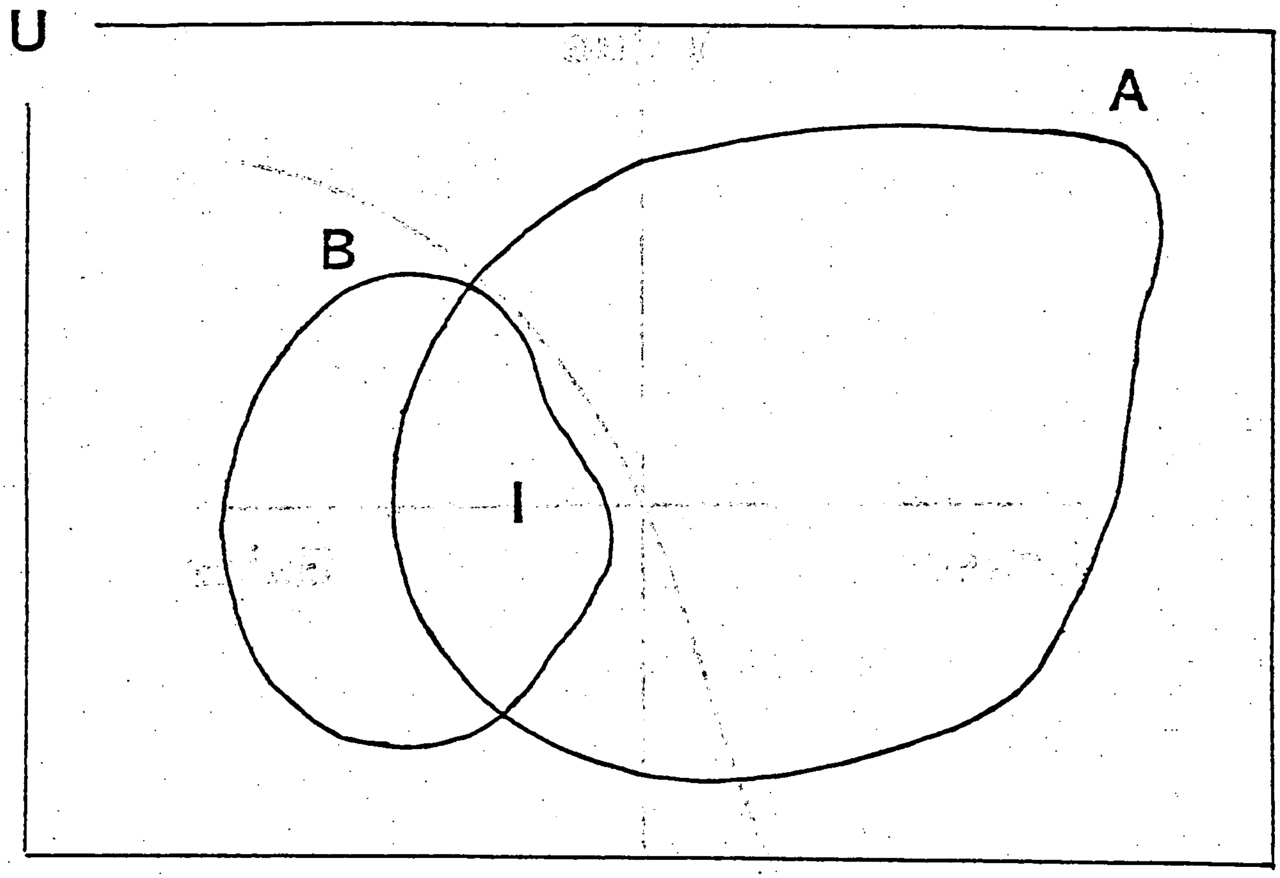

DUPLEX GAMBLE
WIN $2
.6
.4
WIN 0
LOSE $2
.4
.6
LOSE 0
PROBABILITY DISTRIBUTION
PROBABILITY
.5
.16
.48
.36
-$2
0
+$2
PAYOFF
STANDARD GAMBLE
WIN $2
.6
.4
LOSE $2
PROBABILITY DISTRIBUTION
PROBABILITY
.5
.4
.6
-$2
0
+$2
PAYOFF

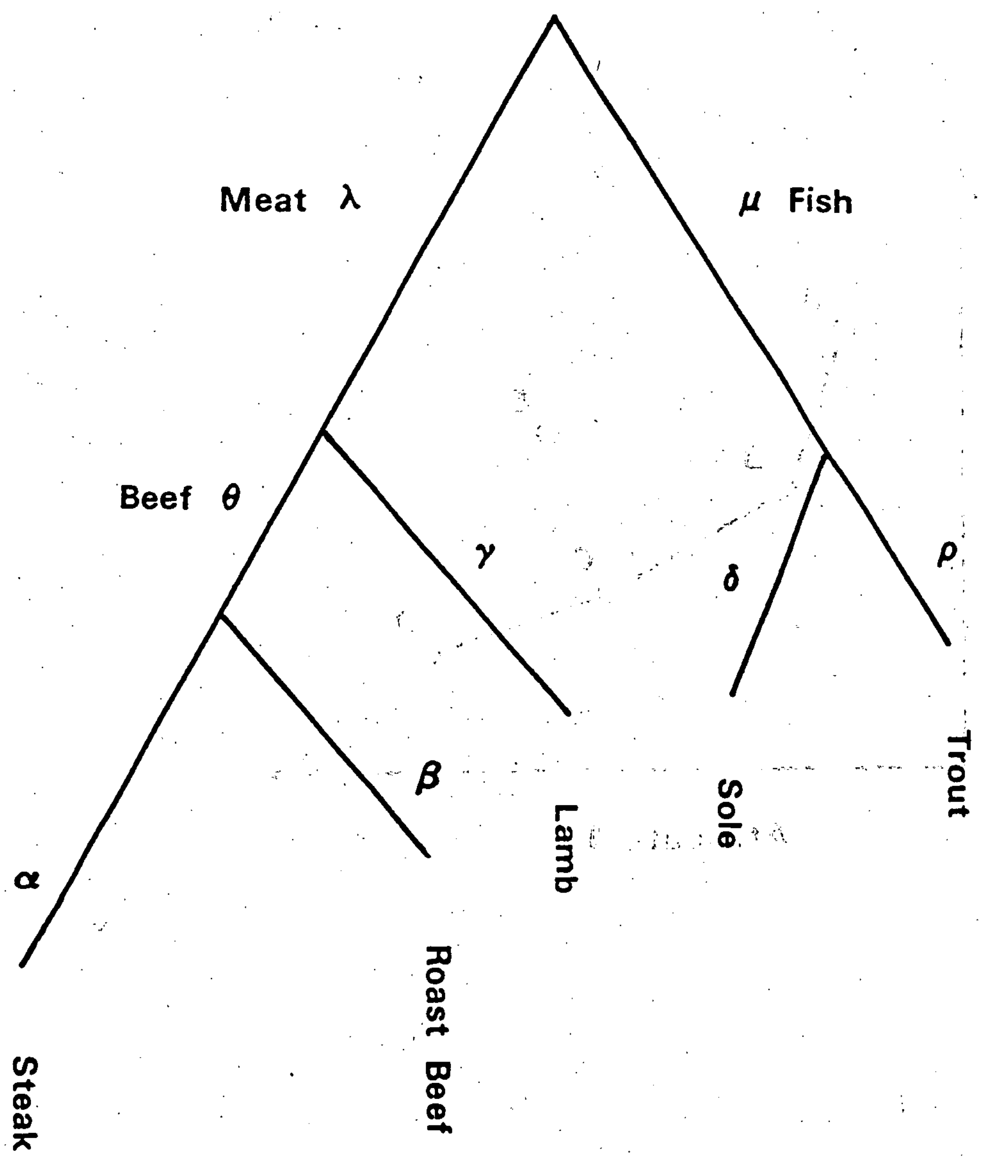

DUPLEX GAMBLE A
WIN 40¢
.4
.6
WIN 0
LOSE 40¢
.7
.3
LOSE 0
PROBABILITY DISTRIBUTION
PROBABILITY
.5
.42
.46
.12
-40¢
0
+40¢
PAYOFF
DUPLEX GAMBLE B
WIN 40¢
.3
.7
WIN 0
LOSE 40¢
.6
.4
LOSE 0
PROBABILITY DISTRIBUTION
PROBABILITY
.5
.42
.46
.12
-40¢
0
+40¢
PAYOFF



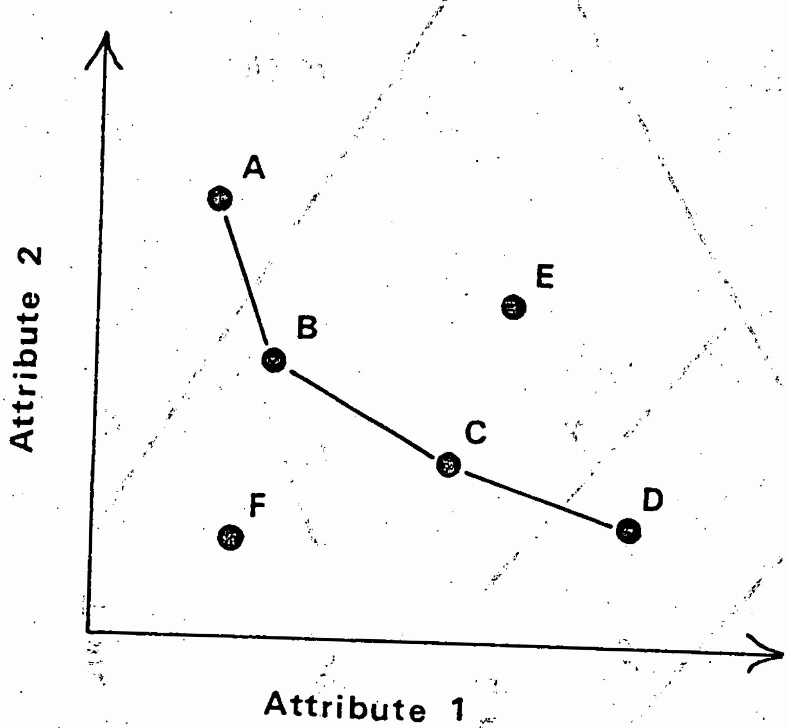
1.0
.75
.50
.25
0
Decision Weight
0
.25
.50
.75
1.0
Probability


Value
Losses
Gains

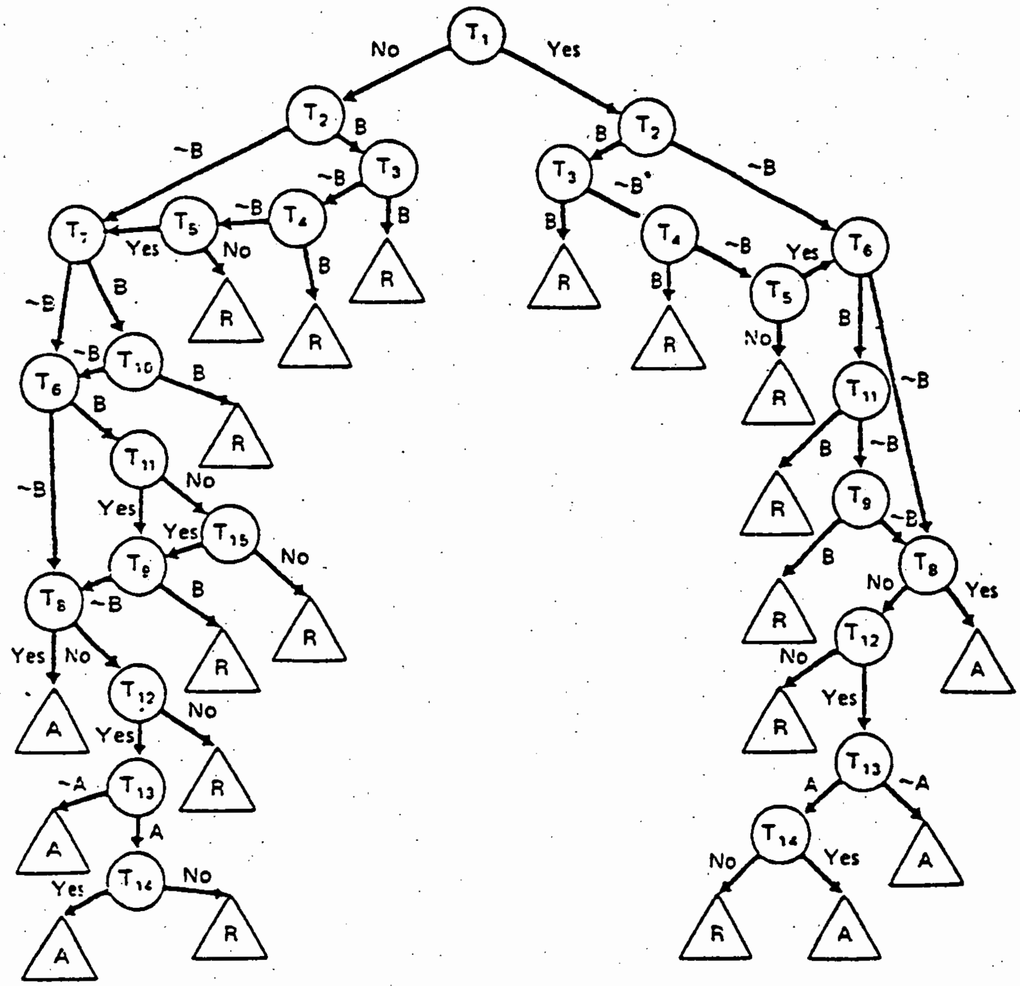

U
A
B
I



Meat λ
μ Fish
Beef θ
γ
δ
ρ
β
α
Steak
Roast Beef
Lamb
Sole
Trout


A
B
C
D
E
F
Attribute 2
Attribute 1

Yield portfolios discrimination net



Key

T_1 — Defensive characteristics
T_2 — Dividend yield ⩾ 4%
T_3 — Dividend yield ⩾ 3.5%
T_4 — Mean yield (past)
T_5 — Have we selected a stock with ⩾ 4%?
T_6 — Mean growth in earnings per share
T_7 — Stability of earnings

T_8 — Is forecasted dividend > 0?
T_9 — Mean growth in working capital
T_{10} — Stability of dividend
T_{11} — Are forecasted earnings > 0?
T_{12} — Is forecasted dividend = 0?
T_{13} — (y) on Relative Value List
T_{14} — Is price > 10% below high?
T_{15} — Is industry depressed-marked "hold"?

B — "Below" A — "Above"
~B — "Not below" ~A — "Not above"

A — Accept R — Reject

Strategies

Raise stakes
Clarify instructions/stimuli
Discourage second-guessing
Use better response modes
Ask fewer questions
Demonstrate alternative goal
Demonstrate semantic disagreement
Demonstrate impossibility of task
Demonstrate overlooked distinction

of problem
be problem
personalized feedback
ensively
em
their responses

explicit
ant information
n
situations
ulations
erts

, 1982a

Table ~~2~~ 1

Matrix Form of a Simple Decision Problem

			State of nature	
			sun (E_1)	rain (E_2)
Alternatives	A_1	carry umbrella	stay dry carrying umbrella	stay dry carrying umbrella
	A_2	leave umbrella	dry and unburdened	wet and unburdened

Table 3
Gambles Used as Stimuli
by Tversky (1969)

Gamble	Probability of Winning	Payoff (in $)	Expected Value
a	7/24	5.00	1.46
b	8/24	4.75	1.58
c	9/24	4.50	1.69
d	10/24	4.25	1.77
e	11/24	4.00	1.83

Table 1. Matrix form of a simple decision problem

Alternative act		States of nature: Sun (θ_1)	States of nature: Rain (θ_2)
A_1	Carry umbrella	(+1) Stay dry, carrying umbrella	(+1) Stay dry, carrying umbrella
A_2	Leave umbrella	(+2) Dry and unencumbered	(0) Wet and unencumbered

Table 1. Matrix form of a simple decision problem

Alternative act		State of nature: Sun (E_1)	State of nature: Rain (E_2)
A_1	Carry umbrella	(+1) Stay dry carrying umbrella	(+1) Stay dry carrying umbrella
A_2	Leave umbrella	(+2) Dry and unburdened	(0) Wet and unburdened