(12)

AD A088698

TECHNICAL REPORT PTR-1091-80-4

THE EXPERIENCED UTILITY OF EXPECTED UTILITY APPROACHES

LEVEL IV

Baruch Fischhoff
DECISION RESEARCH

Bernard Goitein
UNIVERSITY OF MICHIGAN

Zur Shapira
HEBREW UNIVERSITY

April 1980

ARPA

Command & Control
Decision & Forecasting
Systems Program

CYBERNETICS TECHNOLOGY DIVISION
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

9 14 12

TECHNICAL REPORT PTR-1091-80-4

6

THE EXPERIENCED UTILITY OF EXPECTED UTILITY APPROACHES.

12 49

11 Apr 80

by

10

Baruch Fischhoff
~~DECISION RESEARCH~~

Bernard Goitein
~~UNIVERSITY OF MICHIGAN~~

Zur Shapira
~~HEBREW UNIVERSITY~~

411215

Sponsored by

Defense Advanced Research Projects Agency
Under Contract MDA903-80-C-0194
DARPA Order No. 3831

Under Subcontract from

Decisions and Designs, Inc.

15 MDA903-80-C-0194, DARPA Order-3831

April 1980

THE VIEWS AND CONCLUSIONS CONTAINED IN THIS DOCUMENT ARE THOSE OF THE AUTHOR AND SHOULD NOT BE INTERPRETED AS NECESSARILY REPRESENTING THE OFFICIAL POLICIES, EITHER EXPRESSED OR IMPLIED, OF THE DEFENSE ADVANCED RESEARCH PROJECTS AGENCY OR THE UNITED STATES GOVERNMENT.

DECISION RESEARCH
A Branch of Perceptronics
1201 Oak Street
Eugene, Oregon 97401
(503) 485-2400

411215

TABLE OF CONTENTS

SUMMARY

Despite the great progress in devising technological aids to decision makers, most decisions are largely or entirely the product of intuitive, psychological processes. Many times, aids are not available (e.g., under field conditions); at other times, too many decisions have to be made to rely on more than educated intuition (e.g., those made by an officer or executive during a routine day); even when the stakes involved clearly justify investment in aids, time pressures may make that impossible (e.g., crisis situations). Moreover, the use of formal aids has a strong judgmental component.

If these psychological processes are to be improved, they must first be understood. For many years, the dominant model for describing many kinds of decisions has been a weighted average model. In this view, people make decisions on the basis of arithmetic calculations that systematically consider the magnitude of all possible consequences of a decision and the chances of achieving them.

Reviewing many research results, the present report concludes: (a) the ability of the weighted average models to predict the decisions that people make varies from situation to situation; (b) even where they are successful in predicting decisions, these models may not do a very good job of describing the psychological processes that determined them. That is, their predictive accuracy is in part an artifact of the research techniques used.

Without an understanding of how people make decisions, it is difficult to devise ways of enabling them to make better

decisions. The report concludes with several alternative viewpoints on how people make decisions with the ultimate aim of devising decision aids compatible with decision makers' natural modes of thought.

ACKNOWLEDGMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under contract MDA-903-80-C-0194 [redacted] (ARPA Order No. 3871 [redacted]) under subcontract from Decisions and Designs, Inc. to Perceptronics, Inc.

THE EXPERIENCED UTILITY OF EXPECTED UTILITY APPROACHES

A simple and comprehensive rule for making decisions is the following. List all feasible courses of action. For each action, enumerate all possible consequences. For each consequence, assess the attractiveness or aversiveness of its occurrence, as well as the probability that it will be incurred should the action be taken. Compute the expected worth of each consequence by multiplying its worth by its probability of occurrence. The expected worth of an action is the sum of the expected worths of all possible consequences. Once the calculations are completed, choose the action with the greatest expected worth.

Studying the prescriptive and descriptive validity of this model is the preoccupation of researchers in a diffuse field known as behavioral decision theory (BDT). Some look at the model as a whole, others concentrate on its components (e.g., assessments of worth or probability), and still others concern themselves with special cases. The most general form is known as the subjective expected utility (SEU) model. In it, probabilities are treated as being subjective whereas worth is expressed in utility, a generalized measure of desirability. The S is dropped if probabilities are considered to be "objective" (a distinction clarified below). The U, for utility, becomes a V, for value, if an absolute standard of worth, like dollars, is used. The E is dropped if probabilities are ignored, i.e., if there is no element of expectancy and receipt of all consequences is viewed as a certainty. Most of the expectancy-value (E-V) models described elsewhere in this book would in this jargon be called either EU or SEU models; they use subjective judgments of worth and consider uncertainty, but take no position on the meaning of probability.

The essay that follows traces the history of SEU research, with particular emphasis on its varied intellectual roots and the insights and blind spots they have entailed. Like any other discipline, its progress has been speckled by fits, starts and occasional extended pursuit of deadends. These are discussed in moderately blunt terms with an eye to highlighting the hard-earned lessons that seem to be most relevant to expectancy-value (E-V) research. A final section considers less obliquely the question of what the two approaches, E-V and BDT, may have to learn from one another. To presage that discussion, the great disadvantage lies in the complexity of those settings; their great disadvantage lies in the complexity of those settings. Many of the conclusions of behavioral decision theory regarding the intricacies of decision behavior and its study were slowly and painstakingly discovered and documented in highly constrained experimental stituations. Such subtle signals might never have been detected in the flesh. Conversely, field research of the E-V type reminds us of the key variables that should be presented in laboratory work.

FORMAL MODELS OF DECISION MAKING

Attempts to describe how people make decisions took their initial marching orders from prescriptive models, telling how people should make decisions. Variants of the SEU model, as elaborated by statistical and Bayesian decision theory, were the most prominent. These models came from the vicinity of the intersection of economics and philosophy and, indeed, the first experimental studies of decision making were conducted by members of those professions. Two essays by Ward Edwards (1954, 1961) provide comprehensive and stimulating introductions to initial stages of this field.

The appeal of the models was obvious. They reduced the universe of decisions to a common set of primitives (options, probabilities, utilities) about which one could hope to derive universal truths (e.g., people exaggerate low probabilities) in due time. To some extent, this promise has been borne out by reality. Three hundred articles in behavioral decision theory written yearly probably produce more cumulative knowledge than the three thousand publications in personality research (Goldberg, 1974). A second enticement was the presence of an axiomatic basis for these models. Any individual who subscribes to a small set of reasonable rules (e.g., transitivity) should behave in accordance with the model. Prescriptive reasonableness is presumably a necessary condition for the descriptive validity of a law of conscious behavior. In mainstream American economics, the idea that people maximize V or U or EV or SEU or multiattribute SEU has the status of a meta-theory. It goes without saying that such maximization is what people try and manage to do. The goal of the theorist is not to test this assumption, but to divine just what it is that people are maximizing. Like other metatheories, SEU is not falsifiable and thus not a theory at all from this perspective. With sufficient ingenuity, one can always find something that a particular decision maker has maximized in a particular situation.

The story of behavioral decision theory has been the growing realization that SEU often does not describe the decision-making process either as it is designed or executed. The dramatic tension has been provided by SEU's remarkable ability to hang on despite mounting doubts about its descriptive competence. The climax of this yet-incomplete tale seems to have something to do with realizing that a clearly erroneous model still may serve some useful role. The denouement (or perhaps sequel) may involve elaborating and circumscribing that role.[1]

We believe that key clues to the future of the SEU model (and its E-V counterpart) can be found by going back and understanding how it has managed to hang on despite decades of withering attack. The retrospective look that follows suggests the secret to its longevity lies in the fact that the model (a) expresses some fundamental truth in some situations, (b) has a remarkable ability to be close to the mark in situations where its underlying logic is clearly inappropriate and (c) has proponents of heroic tenacity.

Model Fitting

One way to use the SEU model is in explaining past decisions. A set of similar decisions is collected and the researcher sets about identifying a set of predictors (probabilities and utilities) which, when combined according to the SEU rule will account for those decisions. This search may involve not only trying different predictors, but also different ways of measuring and transforming those predictors.

Underlying this approach is a fairly reasonable research strategy. It is extraordinarily difficult to study simultaneously the form of the decision models that people use and the substantive considerations incorporated in them;[2] therefore, let us assume the truth of this general model and devote our efforts to seeing what inputs will make it work. Specifically, assume that people use the SEU model and find what their probabilities and utilities are. With such a strategy, it is hard to lose in the short run. The clever researcher can usually find some set of values in a particular situation that will give good enough predictions or postdictions to maintain faith in the model.

Any descriptive failure by an SEU model can be explained away as a failure in measurement rather than the use of the wrong combination rule. The wrong consequences were included; their worth was assessed in an unreliable or confusing manner; uncertainties were ignored or elicited improperly and so on. Some of the more imaginative fudge factors fall into the categories of transaction costs. Decision making itself is seen as incurring such costs and the process is managed so as to keep those costs down. Thus a decision that seems to select a suboptimal alternative can be reinterpreted as the result of truncating or simplifying the model so as to minimize transaction costs.[3]

The complementary research strategy is to presume that one knows the probabilities and utilities to which people attend and then experiment with different combination rules until one is found that works. The price one pays for strict adherence to either strategy is that as the number of predictors (different variables, measurement techniques, data transformations, etc.) increases, the model becomes increasingly vague and imprecise. Given a sufficiently large set of probabilities and utilities, one can, of course, devise a rule "predicting" past decisions to any desired level of proficiency. In regression terms, by expanding the set of independent variables one can always find a set of predictors (or even one predictor) with any desired correlation with the independent variable. The price one pays for overfitting is shrinkage, failure of the rule to "work" on a new sample of cases. With well-defined problems, the predictor/sample ratio allows one to "correct for shrinkage" and estimate the predictive validity of an SEU model for future decisions. Such corrections are impossible when it is unclear just how many variants of the model have been tried and discarded before the best-fitting one was identified. Thus, when an SEU model "works" on a set of data, it is difficult to know what that proves. Have we proven that the decision makers we are studying are using SEU, and here are their weights? Or has our fishing expedition finally stumbled upon an SEU-like model which mimics their actual behavior?

Simple Models or Simple Processes?

Growing unease with the elusiveness of the proper SEU model was compounded by results emerging from the study of clinical judgment, another intellectual tradition upon which behavioral decision theory draws. Clinical judgment is exercised by a radiologist who sorts X rays of ulcers into "benign" and "malignant," by a personnel officer who chooses the best applicant from a set of candidates, by a crisis center counselor who decides which callers threatening suicide are serious. In each of these examples, the diagnosis involves making a decision on the basis of a set of cues or attributes. When, as in these examples, the decision is repetitive and all cases can be characterized by the same cues, it is possible to model the judges' decision-making policy statistically. Following the logic of the SEU modeling described in the previous section, a set of decisions are gathered and multiple regression is used to determine the cue-weighting scheme that best accounts for the decisions actually made by the judge.

Two decades of such policy-capturing studies persistently produced a disturbing pair of conclusions: simple linear models, using a weighted sum of the cues, did an excellent job of postdicting judges' decisions, despite the judges' claim that they were using much more complicated strategies (Goldberg, 1968, 1970; Slovic & Lichtenstein, 1971). A commonly asserted form of complexity is called "configural" judgment, in which the diagnostic meaning of one cue depends upon the meaning of other cues (e.g., "that tone of voice makes me think 'not suicidal' unless the call comes at mid-day").

Two reasons for this contrast have emerged, each with negative implications for the descriptive validity of SEU. One reason, fed by other developments in cognitive psychology (e.g., Miller, 1956; Simon, 1957), involved the growing realization that combining enormous amounts of information in one's head overwhelms the computational capacity of anyone but an idiot savant. A judge trying to implement a complex strategy simply would not be able to do so with great consistency. Indeed, it is difficult to learn and use even a non-configural, weighted-sum decision rule when there are many cues or unusual relationships between the cues and predicted variable (Slovic, 1974). Since the SEU model is such a weighted-sum rule, these results suggest that people could not use it even if they tried.

The second realization that emerged from clinical judgment research is that simple linear models are extraordinarily powerful predictors. If one can identify and reliably measure the attributes relevant to a decision maker, one can mimic his or her decisions to a large degree with simple models bearing no resemblance to the decision makers' cognitive processes. That is, one can misspecify weights and even combination rules and still do a pretty good job of predicting decisions under very general conditions (Dawes, 1979).

This discovery proves devastating from two perspectives. One is that whatever people are doing will look like the application of an additive linear model (like SEU or E-V). In Hoffman's (1960) term, such models are paramorphic in that they can replicate the input-output relations of the phenomena they are meant to describe without any guarantee of fidelity to the underlying processes. Thus, even if such a model predicts subjects' behavior, one cannot be certain that they are actually using such a decision rule.

Secondly, one cannot take the results of attempts to characterize decision makers' specific policies seriously. If one assumes that people are using an SEU or other additive linear model, their personal decision model could seemingly be captured by finding the weighting schemes that best predict their actual behavior. Unfortunately, the best predicting weights determined, say, by standard regression procedures, become increasingly unstable and uninterpretable when, as is usually the case, there is any dependence between the cues, or multicollinearity. Thus, even positing the (arguable) accuracy of the additive linear model, it is hard to tell what the judge is up to, even when behavior can be predicted quite well.

A possible solution to this problem is to create stimuli or tasks with no dependence between attributes. Slovic (1969), for example, used a factorial design to describe stocks that were evaluated by experienced investors. The variance explained by each dimension (or attribute) could be used to determine its importance in an additive linear decision model. A drawback to this solution is that it goes against another of the intellectual roots of behavioral decision theiry, Brunswik's probabilistic functionalism (1952; Hammond, 1966). From this perspective, the study of behavior involves understanding how people adapt to an uncertain or probabilistic world. Central to their adaptation is learning the natural relationships between cues. A research design that destroys these relationships (e.g., the factorial one described above) lacks ecological validity. When confronted with such an unnatural stimulus environment, the individual can only effect some sort of (meaningless?) adaptation of natural behavior to this unique situation. That compromise may be distinctly unenlightening about real-world behavior, the world which E-V researchers wish to study.

The cumulative effect of these problems was to cast further doubt on the usefulness of research either testing the validity of the SEU model or attempting to explicate its usage through policy capturing. Despite the predictive validity of these attempts, they seemed to yield little trustworthy knowledge about how people actually make decisions. The response of the research community was a parting of ways between two camps which might be described as involving cognitive applied and applied cognitive psychologists (Baddeley, 1977; Wright, 1978). The latter were interested in how people think in the general applied context of decision making; their progress is described in subsequent sections.

The former were interested in specific applied problems whose locus was, at least in part, in cognitive decision making. For them, the predictive power of these models provided a highly useful tool. Often one doesn't care how decisions are made, as long as one can predict their outcome. If it works, a paramorphic model may be good enough for designing an effective marketing campaign or remuneration scheme. Such circumscribed successes are all that many applied E-V researchers need. Although even there, some understanding of how people process information might help present options in the most accessible manner possible.

At times, a valid predictive model might be used to replace clinical judges by formulae modeling their behavior. Called "bootstrapping," this approach calls for having decision makers identify the variables upon which they base their decisions and then "capture" the policy they use with a set of trial cases. That policy, as embodied in a formula, is then used in place of the judges. Even if it does not embody their thought processes, it may mimic their decisions with greater reliability (and hence validity) than the judges themselves. Formulae never have off-days or suffer from fatigue or distraction by irrelevant cues (Slovic & Lichtenstein, 1971). Clearly, people know more than is included in such formulae. But there is no evidence to date indicating that they can convert this sensitivity to the richness of life situations into superior predictions.

Empirically discovering an analytical result by Wilks (1938), Dawes and Corrigan (1974) showed that considerable predictive success is possible without almost any modeling at all. All one has to do is to identify the variables (or attributes) to which a decision maker attends and decide whether they are positively or negatively related to the decision criterion. If these variables are expressed in standard units, they can be given unit weights (+1 or -1, as appropriate). Such a unit weighted model will, under very general conditions, predict decisions as well as a weighted-sum model using regression weights.

Thus, a simple substantive theory indicating what variables people care about when making decisions may be all one needs to make pretty good predictions of their behavior. If they like more of good things and less of bad things, just count up the number of good things an option leads to and subtract the number of bad things and you have a good idea of how favorably it will be viewed. The predictive power of such a simple model provides a base line against which more sophisticated models could be compared.

Obviously, some goods (positive attributes) are more important than others. Therefore, a model using importance weights should, in principle, predict better than one using unit weights. Similarly, goods obtained with high probability should be valued more than those obtained with low probability. Therefore, a model using probability weights should in principle predict better than one using unit weights. However, any unreliability or misspecification of those weights, due to poor procedure or multicollinearity, reduces their contribution very rapidly. In the extreme, models using poorly conceived or executed weighting schemes may succeed in spite of rather than because of their differences from the unit-weight model.

All of these discoveries from the area of clinical judgment have rather chilling implications for the use of modeling techniques for determining whether people use E-V decision-making models and, if so, how. In short, the success of an E-V-type model in predicting behavior proves very little besides membership of the model in the family of powerful linear models. Neither the degree of success, nor the specific importance weights used allow unambiguous interpretation.

Divide and Conquer

Feeling that the SEU model was impregnable and impenetrable when dealt with as a whole, the applied cognitive psychologists among behavioral decision theorists have largely turned to an examination of the model's parts and their validity. Do people accept the axioms upon which the model is based? Are they capable of providing the inputs that it requires? Are their intuitions attuned to the kind of thinking embodied in SEU? In simplified situations, which relieve their computational load, do people exhibit SEU-like behavior? Are they sensitive to factors that have no representation in SEU?

PRESCRIPTIVE VALIDITY

One great attraction of SEU approaches is that they represent reasonable decision rules, ways in which people might want to make decisions; such an attractive prescriptive model would be a sensible point of departure for developing a valid descriptive model. The form of the SEU model ($\Sigma p_i u_i$) states that actions become more attractive as their good consequences become more appealing (u_i increases) and more likely (p_i increases) or as their less appealing consequences become less likely. Since the expected utilities of various possible consequences are added, low or negative utility associated with one consequence can, in principle, be compensated for by sufficiently high utility on others. If people wish to have their decision making guided by these rules, the study of decision making becomes an analysis of their ability to carry them out.

When might people reject these rules?

(a) When more of a good thing is not better. Such situations can, in principle, be handled by having a non-linear or even non-monotonic relationship between the magnitude of a consequence and its utility.

(b) When the appeal of a positive consequence does not increase linearly with the certainty of its being attained (p_i). Atkinson and Feather (1966) and others have shown that at times, people will forego an increment in probability of success in order to be challenged by a situation or to have some suspense associated with its outcome. For example, one might prefer a class in which the top 80% receive A's to one in which A's are given to everyone automatically. SEU models cannot directly accommodate assigning intrinsic values to probabilities of success. However, it is possible to redefine the consequences so as to save the model. In the example, the "cinch A" and the "earned A" are different consequences whose utilities may differ by more than the difference in probabilities (.8 and 1.0), whose values are treated as absolutes.

Because SEU researchers have been interested in developing a general descriptive theory of decision making, there have been relatively few attempts to describe the substantive situations in which such dependencies arise. One does not even know if they occur commonly or merely in a small, but important class of achievement-related circumstances.

One interpretation of such dependence of value upon probability is as a reflection of risk aversiveness. Risk-averse individuals attach a value to certainty itself, whereas risk-seeking individuals do not. There have been some attempts to study individual differences in propensity to take risks, however, these have foundered on the apparent absence of such differences (e.g., Davidshofer, 1976; Wright & Phillips, in press). Poorly understood situational variables seem to overwhelm individual differences in determining risk aversion. Within the achievement literature, the individual difference variable of nAch is well known. It, too, however, is often overwhelmed by situational manipulations of achievement imagery.[4]

(c) When options are evaluated by non-compensatory criteria. Two possible non-compensatory rules are the conjunctive and disjunctive. By the conjunctive rule, an option has to score fairly high on each consequence to be considered. For example, a vacation option must be reasonbly priced, available when needed, suitably sunny and fairly quiet. If an option failed to pass muster on any one of these attributes, its rating on the others would be immaterial, e.g., no amount of sun will compensate for a lot of noise. These minimal levels are, in a sense, non-negotiable demands. According to the disjunctive rule, an option that is adequate on any one attribute is acceptable. For example, an investment opportunity might be chosen if it were good enough as a speculation, tax shelter or hedge against inflation, no matter how badly it rated on the other dimensions. Investment portfolios often include varied items chosen by the same disjunctive rule. Compound strategies are also possible, such as using a disjunctive or conjunctive rule to reduce a large set of options to a smaller one to which a compensatory rule is then applied (Svenson, 1979).

Either partial or total reliance on a non-compensatory rule could spell serious difficulties for the prescriptive (and hence descriptive) validity of SEU models. Einhorn (1971) provided one of the first such demonstrations, although the power of linear models is such that even where they were inferior to non-compensatory models, they still provided relatively good predictions. Lichtenstein, Slovic and Zink (1969) tried to convince people to abandon non-compensatory strategies in favor of using expected value as a decision rule, but to no avail. Choice theory offers a variety of other non-compensatory decision rules (e.g., minimax) whose use has been observed in one situation or another (e.g., Coombs, Dawes & Tversky, 1970). Here, too, our ability to characterize decision situations and the strategies they induce is limited. Attempts to predict the usage and non-usage of compensatory strategies have a distinctly ad hoc character.

(d) When consequences are not evaluated independently. Literal usage of an SEU model requires one to evaluate each consequence attribute of each option by itself and then combine the results. In some situations, however, the decision maker might want to evaluate a particular consequence differently depending upon the value of other consequences. For example, one might like either a slightly uncomfortable chair or a moderate level of ambient noise when deciding where to sit in a lecture, knowing that either will help overcome a tendency to fall asleep. However, a seat that is both noisy and hard will be undesirable since it diverts too much attention to overcoming discomfort. In other words, worth is an interactive function of the attributes.

The concern of E-V researchers over the effects of extrinsic reinforcements on intrinsic motivation (Deci, 1975) reflects another such interaction, as it suggests that the value of intrinsic rewards may be affected by the specific extrinsic reward received. This can, in principle, be handled in SEU models, by evaluating the conjunction of attributes rather than each one separately. In the previous example, one would evaluate separately a noisy and soft seat, a noisy and hard seat, a quiet and soft seat and a quiet and hard seat. With all but the simplest of attribute-attribute interactions, however, this would prove quite laborious in practice.

The prevalence of such interactive evaluations is the most extensively studied aspect of the search for configural judges mentioned earlier. To repeat the results of that search: while clinical judges often believe that they are using configural rules, evidence of consistent configurality is meager. However, it is still unclear whether this contradiction reflects the failure of these judges' introspection (i.e., they don't know what they're doing) or the power of additive linear models.

(e) When options are not evaluated simultaneously. Simon's (1957) notion of "satisficing" grew out of observing the predictive failures of the SEU-like models of classical economics. Satisficers look for decision alternatives that are good enough. In this process, they may use one of the non-compensatory rules described above or a compensatory rule, looking for adequate overall performance. In any case, their search terminates when a satisfactory option has been identified and evaluated. This option may, however, be inferior (in an SEU sense) to other options that are not considered.

An alternative approach to assessing the prescriptive reasonableness of the SEU model is to see whether people accept the axioms from which it is derived. In 1969, Tversky found that while people like to be transitive in their choices (thus fulfilling one axiom), it is possible to design situations in which they are both intransitive and unable to resolve the inconsistency in their choices. He speculated that marketing approaches may be designed to exploit such difficulties. Another demonstration of inconsistency appeared in Zagorski (1975), who showed people pairs of gambles (A, B) and asked them to judge the amount of money V(A-B) that would induce them to trade the better gamble (A) for the worse gamble (B). He demonstrated that one can construct quadruples of gambles A, B, C, and D such that $V(A-B) + V(B-C) \neq V(A-D) + V(D-C)$. In other words, path independence was violated. The difference between gambles A and C depends on whether the intermediate gamble is B or D. The Allais and Ellsberg paradoxes (see Edwards, 1954, 1961) are two demonstrations of people rejecting Savage's independence axiom, according to which preferences between alternatives should be independent of any consequences they have in common.

Until recently, however, few theorists were convinced by these examples. By way of counterattack, MacCrimmon (1968) showed that business executives who violated various axioms could easily be led, via discussion, to see the error of their ways. However, Slovic and Tversky (1974) challenged MacCrimmon's discussion procedure on the grounds that it pressured subjects to accept the axioms. They presented subjects with arguments for and against the independence axiom 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 maximize its acceptability, yet still found that the independence axiom was rejected. Even MacCrimmon's faith in many of the key axioms has been shaken by recent data, leading him to suggest that reevaluation of the theory is in order (MacCrimmon & Larsson, 1976).

To be most useful, such a reevaluation would have to go beyond demonstrating that violation or rejection of the axioms is possible. It would have to give some guidelines as to the prevalence and distribution of violations. Are they just concocted curiosities? Or do they represent modal behavior in some important realms?

Kahneman and Tversky (1979) attempted this sort of reevaluation, presenting evidence for two pervasive violations of SEU theory. One, the "certainty effect," causes consequences, both positive and negative, that are obtained with certainty to be given more weight than uncertain consequences. The Allais paradox may be due to this effect. The second, labeled the "reference effect," leads people to evaluate alternatives relative to a reference point corresponding to their present status, expectation or adaptation level (Helson, 1947, 1959). By altering the reference point, formally equivalent versions of the same decision problem may elicit different preferences. These effects pose serious problems for the validity (prescriptive and descriptive) of SEU approaches. The power of Kahneman and Tversky's theory is its ability to predict responses to decision problems posed in particular ways and, in particular, to predict behavior that is contrary to predictions of the SEU model. It needs, however, to be complemented by a substantive theory regarding the way in which decision questions are posed by nature and interpreted by observers.

AVAILABILITY OF INPUTS

Let us restrict our attention to situations in which people might wish to engage in SEU-like behavior. Doing so means providing (and eventually integrating) lists of feasible options and possible consequences along with assessments of their probabilities and values. One needn't perform these tasks well in order to engage in SEU-like behavior. Poor performance will, however, lead to suboptimal decisions. As the lists become more and more incomplete and the quality of the assessments deteriorates, the resultant decision will tend to deviate from that obtained by the best possible usage of the SEU model.

Thus making decisions not in one's own best interests (as defined by the "best possible" SEU decision) need not mean that the SEU model is an invalid descriptor of what people are doing. Nonetheless, acute inability to provide the inputs required to use the SEU model well must cast some doubts on the extent of its usage. At some point, one must stop and ask, would people persist in doing something they do so poorly? To take an analogous problem, I may realize the essential wisdom of using the 1040 Form for income tax and itemizing my deductions in order to reduce my obligation. If, however, experience shows that I typically do it wrong and land in trouble, I may resort to less complicated decision rules. Whether or not it is reasonable to expect people to persist at suboptimal SEU behavior, rather than opting for some simpler rule, requires consideration of two questions: (a) Just how good or bad are they? and (b) How likely are they to learn about their mistakes? If people both could not perform these tasks and realized their limitations, the SEU model would be a less likely candidate for how people try to make decisions.

What else could people do, besides trying to list and assess the expected consequences of all courses of action open to them? For one, they could try not to think at all, but rely on non-analytical decision rules like "this is (most like) what I've always done" or "this is what my (most expert) friends tell me to do" or "this is what everyone else is doing." Or they could refuse to make one final decision, preferring to muddle through by trial and error, making small incremental decisions that afford an opportunity to change courses if things aren't going well.

Relying on the collective wisdom of one's own experience (habit) or of one's peers' experience (tradition) or of one's peers' current feelings are all ways of externalizing the problem. Each recruits a number of people to think about issues too complex and poorly understood to analyze alone. Both common wisdom and such au courant techniques as Delphi advocate such thought sharing. One could, of course, consult these sources for opinions that would eventually be used in a personal SEU model. Or one could reject the methodological individualism of SEU decision making in favor of whatever conclusion emerges from these chaotic group processes. Such reliance on group rather than individual "rationality" is well documented in work situations (Salancik & Pfeffer, 1977, 1978; White & Mitchell, 1979). Perhaps the best known example is the emergence of group production norms in defiance of piece-work pay schemes.[5]

One of the few exercises in complex analytic decision making imposed upon most people is completion of income tax forms. The success of H & R Block suggests the reluctance of the rest of us to face the challenge.

Problem Structuring

Like other analytic approaches, SEU begins with a structuring of the problem, namely listing all relevant options and consequences. The obvious performance standard here is completeness. A modest amount of systematic and anecdotal evidence suggests that people have difficulty independently producing complete problem representations. When asked to judge completeness of problems presented by others, however, they do not seem to be very sensitive to these inadequacies. What is out of sight is effectively out of mind (Fischhoff, Slovic & Lichtenstein, 1977, 1978). Thus in the short run at least, people may not realize this limit on their analytical decision-making ability. In the long run, omitting options and consequences should lead to poor decisions. However, life seldom sends us large enough batches of unambiguous signals to make it clear just how good our decision making is and where the problems lie (Fischhoff, 1975; Einhorn & Hogarth, 1978). While unrecognized incompleteness seems to be quite general, there presumably are decision problems whose most important features are readily uncovered by people's intuitive procedures. Identifying that set of problems would require the sort of substantive theory of people in situations which has interested E-V researchers.

Assessing Probabilities

A cornerstone of SEU thinking is that we all live in an uncertain world. According to the Bayesian or subjectivist position that uncertainty reflects the limits of our understanding. From this position, all statements are implicitly or explicitly qualified by our degree of belief in them. Degrees of belief are numerically expressed in subjective probabilities, i.e., expressions with the form "My personal probability that Statement A is true is .XX". Such probabilities are entirely in the eye of the beholder and are not properties of the world. Two observers of a situation can in principle assign different probabilities of being true to the same statement about it, either because they interpret what they see differently or have different background information.

In fact, the whole notion of "objective probability" is confused. Assigning any probability requires some interpretation of the observed situation and that act of interpretation imputes a subjective element. What are commonly thought of as objective probabilities merely refer to situations in which there is consensus among reasonable observers about how to interpret existing evidence. The probability of a fair coin falling on "heads" on its next flip is commonly held to be .5, but that is only because observers agree on how to evaluate past performance and relate it to subsequent performance (e.g., what constitutes an adequate series of independent, equivalent trials). Bayesians do not talk about <u>estimating</u> probabilities which implies that there is something "out there" in the world to be appraised, but of <u>assessing</u> probabilities, signifying the evaluation of an internal state.

If probabilities are subjective, does it make any sense to consider their validity? Individual probabilities can almost never be evaluated (except in the case where one says "There is <u>absolutely</u> no chance of this statement being wrong" regarding a statement that is, in fact, wrong).

Sets of subjective probabilities may, however, be evaluated according to two criteria. One is consistency with the laws of probability or (internal) coherence (Kyburg & Smokler, 1964). For example, the probabilities assigned to a statement and its complement should total 1.0. The second is "calibration," applicable when the truth of statements is known to the probability assessor. The well-calibrated probability assessor should have more true statements associated with high than with low probabilities. Specifically, XX% of the statements assigned probability .XX should be perceived to be true. That is, if I am well calibrated, 70% of the time when I say "there is a .7 chance of this statement being true," it should be true.

The earliest studies of probability assessment were restricted to situations in which consensus could be reached on "objective" probability values. These experiments used repetitive series of events (like drawing red or blue balls from an urn) and defined probability as relative frequency of occurrence. Research showed that people were quite good at appraising such frequentistic probabilities (Peterson & Beach, 1967).

More recently, the focus of research has turned to how people assign probabilities to statements, including unique events (e.g., Carter will be re-elected), where relative frequency has no meaning. This shift seems to reflect a growing acceptance of the subjectivist position, a feeling that one should study probabilities in their most general form and a realization that relatively few important events in people's lives are thought of in terms of relative frequency.

Here, performance has been less than outstanding. Experiments using a variety of tasks, response modes and subjects have shown that probability assessments tend to be poorly calibrated (Lichtenstein, Fischhoff & Phillips, 1977). The most common type of error is overconfidence, i.e., people think that they know more than is actually the case. Where observed, this bias seems to be so robust that people are willing to engage in highly disadvantageous bets based upon their confidence judgments (Fischhoff, Slovic & Lichtenstein, 1977).

Various forms of incoherence have also been observed, although without any clear indication of their prevalence. For example, Wyer (1974) found that people tend to exaggerate the probability of the conjunction of two events, relative to the probabilities assigned to the two individually. Kahneman and Tversky (1973) found that rather than combining background (base rate) information with information regarding a particular case, people tend to ignore base-rate probabilities. Slovic, Fischhoff & Lichtenstein (1976) demonstrated situations in which the judged probability of compound events was larger than the probabilities of their constituent events.

It is disturbing that $P(A \cap B)$ should ever be greater than $P(A)$ or $P(B)$, but some notion of prevalence is needed to decide whether the problem is so bad as to indicate either (a) that people would realize their limits and avoid probabilistic thinking or (b) that the quality of people's probabilistic thinking is so poor as to suggest that they seldom engage in it. Such a study of prevalence and extent would have to be accompanied by an analysis of how bad performance would have to be for people to notice. In this context, the folklore of survey research may be instructive. Popular wisdom there holds that one cannot ask for numerical probability assessments from a random sample of the population without getting disturbingly high rates of non-response (i.e., over 20%). Regarding detailed verbal statements of probability, the experience of the 1977 Quality of Employment Survey (Quinn & Staines, 1978) is instructive. Pretest work resulted in probability assessments being gathered on a response scale with only two alternatives: "Yes, it is likely" and "No, it is not likely." The test-retest reliability of more detailed verbal probability assessment has been found at times to be surprisingly low (r = .52-.56 for DeLeo & Pritchard, 1974; Lied & Pritchard, 1976). On the other hand, people do seem to have some appreciation of the meaning of probabilistic weather forecasts (Murphy, Lichtenstein, Fischhoff & Winkler, 1979). Perhaps they feel more comfortable with the probability of rain because the event is highly repetitive and their task is passive (listening).

Perhaps the most disturbing aspect of these results for a proponent of SEU descriptive models is not the presence of biases, but the thought processes that seem to underlie both erroneous and accurate probability assessments. Tversky and Kahneman (1974) have proposed that many probabilistic judgments are produced by using rules of thumb or heuristics whose internal logic bears little resemblance to the rules of probability. These are generally effective ways of coping with an uncertain environment that both deny its probabilistic character and preclude learning about some basic phenomena (like regression to the mean). Often they lead to substantially biased judgments and decisions not in the individual's best interest.[6] To the extent that these heuristics capture people's thought processes, they (a) argue strongly against the notion that we are efficient probabilistic functionalists, well-tuned to the uncertain structure of our environment, and (b) even when people do incorporate uncertainty in their decision making, they do so in terms quite different from the formally defined probabilities appearing in SEU models.

Assessing Values

The study of value judgments and their validity by behavioral decision theory researchers has languished relative to the study of probability judgments. The presumed reason for this apparent disinterest is the absence of an acceptable criterion by which to evaluate value judgments. Such judgments would seem to be the last redoubt of unaided intuition. Who knows better than an individual what he or she prefers?

Recent research has, however, revived and elaborated a lesson long known to attitude researchers. Subtle changes in how value questions are phrased and responses are elicited can have marked effects on the preferences people express. This lability in value judgments seems to have important implications for both how values are conceptualized and how they are studied.

An article of faith among students of value, choice and attitude judgments is that people have reasonably well-defined opinions regarding the desirability of various events. Although these opinions may not be intuitively formulated in numerical (or even verbal) form, careful questioning can elicit judgments representing people's underlying values. From this stance, elicitation procedures are neutral tools, bias-free channels which translate subjective feelings into scientifically usable expressions. They impose no views on respondents, beyond focusing attention on those value issues of interest to the investigator.

It is by no means obvious, however, that people's values are well defined. Observed test-retest reliabilities in the values of desired outcomes (within situations) have been fairly modest (r = .48, .60 in DeLeo & Pritchard, 1974; Lied & Pritchard, 1976) and remarkably little consistency in the rated valence of outcomes across situations was observed by Muchinsky (1977). Lawler (1971) found wide variations in the importance attached to one central outcome of work, pay, at least partially due to variations in elicitation procedures.

What happens in cases where people do not know, or have difficulty appraising, what they want? Under such circumstances, these procedures may become major forces in shaping the values expressed, or apparently expressed, in the judgments they elicit. They can induce random error (by confusing the respondent), systematic error (by hinting at what the "correct" response is), or unduly extreme judgments (by suggesting clarity and coherence of opinion that are not warranted). In such cases, the method becomes the message, a reflection of the researcher's tools rather than the respondent's wishes.

People are most likely to have clear preferences regarding issues that are familiar, simple, and directly experienced.[7] Each of these properties is associated with opportunities for trial-and-error learning, particularly such learning as may be summarized in readily applicable rules or homiles. Those rules provide stereotypic, easily justifiable responses to future questions of value. When adopted by individuals, they may be seen as habits; when adopted by groups, they constitute traditions.

The power of these rules of thumb for assessing values comes from their development and application to simple and repetitive problems. Their viability becomes suspect when the issues are unfamiliar and complex, the old intuitions impotent, the old rules untested and perhaps untestable. Unfortunately, these are precisely those situations in which values are worth studying. In them, however, we may never have considered the implications of the values and beliefs acquired in simpler settings. As a result, we may have no articulated preferences. In some fundamental sense, our values may be incoherent, not thought through. In thinking about what are acceptable levels of risk, for example, we may be unfamiliar with the terms in which issues are formulated (e.g., social discount rates, miniscule probabilities, or megadeaths). We may have contradictory values (e.g., a strong aversion to catastrophic losses of life and a realization that we're not more moved by a plane crash with 500 fatalities than one with 300). We may occupy different roles in life (parents, workers, children) which produce clear-cut, but inconsistent values. We may vacillate between incompatible, but strongly held, positions (e.g., freedom of speech is inviolate, but should be denied to authoritarian movements). We may not even know how to begin thinking about some issues (e.g., the appropriate tradeoff between the opportunity to dye one's hair and a vague, minute increase in the probability of cancer 20 years from now). Our views may undergo changes over time (say, as we near the hour of decision or the consequence itself) and we may not know which view should form the basis of our decision.

Such inarticulated preferences are hardly compatible with the sort of rigorous systematic thinking required by SEU decision making. Furthermore, they give pause for thought about just what subjects are giving us when we elicit the values needed to test the conformity of their behavior with the SEU model.

Listing a few specific effects may indicate the power an elicitor may deliberately, or inadvertently, wield in shaping, distorting or even creating expressed preferences. The desirability of possible outcomes is often evaluated in relation to some reference point. That point could be one's current (asset) position, or an expected level of wealth (what someone with my talents should be worth at time t), or that possessed by another person. Shifts in reference point are fairly easily effected and can lead to appreciable shifts in judged desirability, even to reversals in the order of preference. Consider, for example, how one might think about the same safety program conceptualized in terms of lives saved or lives lost, with the respective reference points of the current situation or an ideal one. As one gets closer to an event with mixed consequences, the aversiveness of its negative aspects may increase more rapidly than the attractiveness of its positive aspects, making it appear, on the whole, less desirable than it did from a distance. People may have opposite orders of preference for gambles when asked which they prefer (which focuses their attention on how likely they are to win) and when asked how much they would pay to play each (which highlights the amount to win). People may prefer to take a chance at losing a large sum of money rather than absorb a smaller sure loss, but change their mind when the sure loss is called an insurance premium. A relatively unimportant attribute may become the decisive factor in choosing between a set of options if they are presented in such a way that that attribute affords the easiest comparison between them.

Three important features of these shifting judgments are: (a) people are typically unaware of the potency of such shifts in their perspective, (b) they often have no guidelines as to which perspective is the appropriate one, and (c) even when there are guidelines, people may not want to give up their own inconsistency (Fischhoff, Slovic & Lichtenstein, 1978, 1980).

Elicitors must decide at some point whether or not they have adequately captured their respondent's values. The usual criteria are reliability and internal consistency (e.g., transitivity). However, where the task is poorly understood because of complexity or unfamiliarity, consistency of response within a given elicitation mode may tell us little beyond the power of that mode to impose a particular perspective. Consistency of response is a necessary but not sufficient condition for coherence of the underlying values. Greater insight into values may come from posing diverse questions in the hope of eliciting inconsistent responses. Therefore, one would want to start the study of values with methodological pluralism and treat inconsistency in expressed values as a success rather than a failure of measurement, for it indicates contexts defined sharply enough to produce a difference.

If one is interested in how people express their values in the real world, one question may be enough. That world often asks only one question (e.g., in a ballot measure). A careful analysis of how an issue is posed may allow one to identify that question and accurately predict responses.

If one is interested in what people really feel about a value issue, there may be no substitute for an interactive, dialectical elicitation procedure, one that acknowledges the elicitor's role in helping the respondent to create and enunciate values. That help would include a conceptual analysis of the problem and of the personal, social, ethical value issues to which the respondent might wish to relate.[8]

DECISION CONTEXTS

While behavioral decision theorists have not developed a theory of decision situations to complement their evolving theory of the individual decision maker, they have upon occasion attempted to replicate their laboratory studies in the real world. These experiments at roughing it have been prompted by (a) the feeling that decision theory like other areas in applied cognitive psychology should work in the world; (b) the refusal of economists to accept (or even look at) laboratory data; (c) the hope of being able to elaborate their theories by embedding them in a broader context and (d) continued interaction with decision analysts, purveyors of SEU as a normative guide to decision making who are particularly attuned to the subjective quality of judgments of fact and value.

One obvious step in the direction of realism is to use real rather than hypothetical stakes in studies of decision making. A second is to use experts rather than naive subjects. Although available evidence is modest, neither manupulation has so far provided sufficiently dramatic results to cast serious doubt on most of the research conclusions cited above

(Slovic, Fischhoff, Lichtenstein, 1979; Tversky & Kahneman, 1974). A third obvious step, using more realistic stimuli, has, it seems, only strengthened conclusions regarding the lability of judgments of value and liabilities of judgments of fact (Lichtenstein et al., 1978; Fischhoff, Slovic & Lichtenstein, 1978).

None of these manipulations, however, leads to the study of actual judgments in situ. Probably the most elaborate excursion of behavioral decision theory into the real world has been a study of disaster insurance protection by Kunreuther et al. (1978). Despite massive investments in flood control projects, flood damage in the U.S. continues to climb. The main reason seems to be the overdevelopment of flood plains caused in part by the flood control projects themselves. By eliminating frequent minor floods, these projects have unduly reduced residents' feeling of flood danger. As a result, when a sufficiently large rain or thaw somes along and overwhelms the project, much more property is exposed. Thus floods are less frequent, but the loss from each is much greater than prior to the project. In order to reduce losses, the U.S. Congress enacted a flood insurance program which mandated land-use planning as the price for making insurance available at highly attractive rates. "Highly attractive" was determined by economists who assumed that residents shared the risk and cost data in the hands of the planners and combined them according to SEU principles. Unfortunately, almost no one bought the insurance.

A national survey of flood-prone areas designed to understand the failure of the program discovered that residents' judgments of risks and costs were very different than those of the experts. Often these misjudgments seemed to have the same roots as the judgmental biases observed in laboratory studies. Obviously, even if residents were SEU decision makers, they could not be expected to behave in the predicted way (i.e., purchase insurance) if they did not share the planners' information base. The design of Kunreuther's study allowed an appraisal of how favorable insurance should be to these individuals given what they thought to be the facts of the matter and assuming that they were SEU decision makers. A straightforward derivation translated attractiveness of insurance into an index K, with higher K values being associted with more attractive insurance situations. Comparison of individuals who had and had not purchased insurance revealed no difference at all in the distribution of K values. Thus whatever guided insurance purchase decisions, it was not an SEU analysis based on either the "real" (experts') or "imagined" (residents') view of the situation.

What did guide decisions? Kunreuther's best guess was that people relied upon a series of informal decision rules like the answers to: Are my neighbors and relatives buying it? What does my insurance agent say? Do I have the capacity to worry about one more thing?

The application of E-V theory to predicting effort and performance in the work place may be considered a prolonged extension of SEU theory into the real world. In an extensive review of this effort, Schwab et al. (1979) found that on the average, only 9% of the variance in effort or performance is explained. Granting that their review considered only between-subject applications of this within-person behavioral choice theory (thereby increasing the amount of error variance), analogous within-person studies rarely explain more than 25% of the variance. Furthermore, with one exception, the few factors that Schwab and his colleagues identify as contributing to better prediction are either not those prescribed by E-V models, or antithetical to them. The authors conclude by noting that:

> Despite these qualifications, there is a nagging suspicion that expectancy theory overintellectualizes the cognitive processes people go through when choosing alternative actions (at least insofar as choosing a level of performance or effort is concerned). The results of the present review are consistent with this suspicion (p. 146).

More pessimistically, Staw (1977) suggests that ordinary choice behavior is better modeled as "impulse buying" decisions of consumers than on the lines of SEU theory.

Perhaps more revealing than the contexts to which researchers have attempted to extend laboratory results are the contexts that they have avoided. Situations to which they have assumed that results will not generalize may provide an implicit partial theory of the environment. One characteristic of most behavioral decision theory experiments is that every attempt is made to eliminate any temptation for strategic responding. Rather, subjects are rewarded for responding as honestly as they can. However, there obviously are situations where it may pay to lie (e.g., when the boss asks me about my work environment) and they may engender rather different reporting behavior. A second characteristic is the absence of time and emotional pressure creating the sort of "hot cognition" studied by Janis and Mann(1977). A third characteristic is isolation of the individual from social interactions, which may deprive the individual of needed decision-making aids (i.e., talking to others) and foster individualism. A fourth characteristic of most studies is the preclusion of a no-choice or procrastination option (Corbin, 1980).

Each of these characteristics might in principle represent an environmental variable restricting the validity of behavioral decision theory results. Or each might provoke minor perturbations like the use of real stakes or the shift from naive to expert subjects.

CONCLUSION

The story of SEU research has in some senses been a tale of deadends and hard-earned lessons. An enormous amount of effort was devoted to "capturing" decision strategies before it was realized that the power of linear models virtually precluded learning what weights individuals were giving to different attributes or if they were using an SEU decision rule at all. Many studies of how people estimate relative frequencies were conducted before researchers realized that whatever their intrinsic interest such tasks were not particularly relevant to the sort of uncertainty in most decision situations. For years, researchers derived satisfaction from the elicitation of consistent, reliable value judgments, before beginning to worry that such orderliness was a product of their methods. Being rational individuals, decision scientists have repeatedly assumed that the inherent reasonableness of the SEU model would make it an acceptable normative guide to decision making. Those who doubted the model have naively felt that showing spot inconsistencies or rejection of axioms would cause the model to keel over.

Having disabused ourselves of these particular illusions, where do we go from here? Obviously, we should have better respect for the power of our methodology to produce orderly but misleading results. Substantively, we should begin to develop an understanding of decision environments which will enable us to understand how general laboratory results are and how far reaching their effects (Ebbesen & Konecni, 1980; Howell & Burnett, 1978). One key component of this understanding will be an error theory explicating the implications of poorly structuring decisions, assessing probabilities and values, and combining the various components. The sensitivity of our decisions and their consequences to deviations from optimality will clarify which principles of decision making need to be and can be learned from experience (Einhorn & Hogarth, 1978; Fischhoff, 1980a, b).

Theoretically, we need to go beyond simply trying to falsify SEU theory which, if only because of the power of linear models, will almost always provide at least mediocre predictions. Instead, we need to move on to sophisticated falsification (Lakatos, 1970), finding theories that do what SEU does and at least a little more. The shape of those theories might be new decision calculi with different primitives and combinations rules, like Kahneman and Tversky's Prospect Theory (1979); or process-tracing models that attempt to treat not only input-output relationships, but also the intervening thought processes (Payne, Braunstein & Carroll, 1978; Svenson, 1979); or contingency-based models that preserve some of the SEU logic but also incorporate a theory of the environment and the way in which it is sequentially decomposed (Beach & Mitchell, 1978); or inventories of the rules of thumb that people use to supplement or replace analytical decision making (Kunreuther et al., 1978); or predictions of how the habits ingrained from making repetitive decisions with an opportunity for trial-and-error learning lead one astray when one must make analytic decisions and get them right the first time.

FOOTNOTES

1. One source of resistance only tangentially relevant to our story lies in the disciplinary prejudices of the economists and (to a lesser extent) the philosophers instrumental in launching SEU. The sophisticated mathematical derivations and elegant symbol manipulations which are the stock in trade of these two groups require a degree of formalism like that found in SEU. The theory did not fall over dead with the inconsistencies identified by Allais or Ellsberg, for example, because neither offered an alternative calculus upon which these professionals could work their magic.

2. Functional measurement is one of the few techniques designed to do this (Anderson, 1978). While studies in this vein have produced useful insights, literal acceptance of their conclusions requires some heady assumptions, particularly regarding the verisimilitude of responses to involved factorial designs and highly schematic stimuli. For example, one must be convinced that subjects do not solve the problems presented by such tasks by concocting situation-specific decision rules that emerge, upon analysis, as highly systematic behavior.

3. See Beach and Mitchell (1978) for a predictive model using transaction costs, albeit not labeled as such, as a major determinant in the selection of decision-making strategy.

4. Influence in the opposite direction is also possible. Probabilities may be influenced by utilities, as when optimists exaggerate the probability of good things happening. There appears at the moment not to be any hard evidence that such distortion occurs (Wallsten, 1971).

5. A resolute SEU devotee could still claim that the group had

been used to restructure the decision problem offered by management (piecework). Adherence to group production norms is actually the result of individual SEU maximizing in the context of the new decision problem which incorporates group action as a new option, group sanctions and collectively attained pay schemes as new consequences and collective bargaining power as a new fact of life affecting the probabilities of all consequences.

6. While Kahneman and Tversky's research has come to be known for its identification of errors, they make no statement about how bad judgment is in general. They focus on errors because there are fewer ways to explain a pattern of errors than a pattern of success and because suboptimal behavior in conditions encouraging optimality suggests deep-seated cognitive tendencies.

7. For example, Ash, Levine & Edgell (1979) found that the reliability of the rated desirability of attributes increases with respondents' direct experience with the job.

8. As an example of the power of the interactive approach to assessing values, Matsui and Ikeda (1976) and Rosenberg (1956) were better able to predict people's decisions when using outcomes generated on the spot by subjects than with a more comprehensive list of outcomes that they, as experimenters, provided.

REFERENCES

Anderson, N. H. Progress in cognitive algebra. In L. Berkowitz (Ed.), _Cognitive theories in social psychology_. New York: Academic Press, 1978.

Ash, R. A., Levine, E. L. & Edgell, S. L. Exploratory study of a matching approach to personnel selection: The impact of ethnicity. _Journal of Applied Psychology_, 1979, _64_, 35-41.

Atkinson, J. & Feather, N. (Eds.). _A theory of achievement motivation_. New York: Wiley, 1966.

Baddeley, A. D. _Applied cognitive and cognitive applied psychology: The case of face recognition_. Paper presented at Uppsala Conference on Memory, June, 1977.

Beach, L. R. & Mitchell, T. R. A contingency model for the selection of decision strategies. _Academy of Management Review_, 1978, _3_, 439-449.

Brunswik, E. _The conceptual framework of psychology_. Chicago, Illinois: University of Chicago Press, 1952.

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

Corbin, R. M. A theory of choice should not be based on choice alone. In T. Wallsten (Ed.), _Cognitive processes in choice and decision behavior._ Hillsdale, N. J.: Erlbaum, 1980.

Davidshover, L. O. Risk-taking and vocational choice: A reevaluation. _Journal of Counseling Psychology_, 1976, _23_, 151-154.

Dawes, R. M. The robust beauty of improper linear models. _American Psychologist_, 1979, _34_, 571-482.

Dawes, R. M. & Corrigan, B. Linear models in decision making. _Psychological Bulletin_, 1974, _81_, 2, 95-106.

Deci, E. L. _Intrinsic motivation_. New York: Plenum, 1975.

DeLeo, P. J. & Pritchard, R. D. An examination of some methodological problems in testing expectancy-value models with survey techniques. Organizational Behavior and Human Performance, 1974, 12, 143-148.

Ebbesen, E. B. & Konecni, V. J. On the external validity of decision making research. In T. Wallsten (Ed.), Cognitive processes in choice and decision making. Hillsdale, N. J.: Erlbaum, 1980.

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

Edwards, W. Behavioral decision theory. Annual Review of Psychology, 1961, 12, 473-498.

Einhorn, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. Organizational Behavior and Human Performance, 1971, 6, 1-27.

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

Fischhoff, B. Clinical decision analysis. Operations Research (issue devoted to decision analysis), 1980a, 28, 28-43.

Fischhoff, B. For those condemned to study the past: Reflections on historical judgment. In R. A. Schweder & D. W. Fiske (Eds.), New directions for methodology of behavior science: Fallible judgment in behavioral research. San Francisco: Jossey-Bass, 1980b.

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

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, 342-355.

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.

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

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

Goldberg, L. R. Objective diagnostic tests and measures. Annual Review of Psychology, 1974, 25, 343-366.

Hammond, K. R. (Ed.), The psychology of Egon Brunswik. New York: Holt, Rinehart & Winston, 1966.

Helson, H. Adaptation level as frame of reference for prediction of psychophysical data. American Journal of Psychology, 1947, 60, 1-29.

Helson, H. Adaptation level theory. In S. Koch (Ed.), Psychology: A study of a science, Vol. 1. New York: McGraw Hill, 1959.

Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 47, 116-131.

Howell, W. C. & Burnett, S. A. Uncertainty measurement: A Cognitive taxonomy. Organizational Behavior and Human Performance, 1978, 22, 45-68.

Janis, I. & Mann, L. Decision making. New York: The Free Press, 1977.

Kahneman, D. & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

Kahneman, D. & Tversky, A. Prospect theory. Econometrica, 1979, 47, 263-291.

Kunreuther, H., Ginsberg, R., Miller, L., Sagi, P., Slovic, P., Borkan, B. & Katz, N. Disaster insurance protection: Public policy lessons. New York: Wiley Interscience, 1978.

Kyburg, H. E. Jr. & Smokler, H. E. Studies in subjective probability. New York: Wiley, 1964.

Lakatos, I. Falsification and the methodology of scientific research programmes. In Lakatos, I. & Musgrave, A. (Eds), Criticism and the growth of knowledge. London: Cambridge University Press, 1970

Lawler, F. E., III. Pay and organizational effectiveness: A psychological view. New York: McGraw Hill, 1971.

Lichtenstein, S., Fischhoff, B. & Phillips, L. D. Calibration of probabilities: The state of the art. In H. Jungermann & G. deZeeuw (Eds.), Decision making and change in human affairs. Amsterdam: D. Reidel, 1977.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M. & Combs, B. Judged frequency of lethal events. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 551-578.

Lichtenstein, S., Slovic, P. & Zink, D. Effect of instruction in expected value on optimality of gambling decisions. Journal of Experimental Psychology, 1969, 79, 236-240.

Lied, T. R. & Pritchard, R. D. Relationships between personality variables and components of the expectancy-valence model. Journal of Applied Psychology, 1976, 61, 463-467.

MacCrimmon, K. R. Descriptive and normative implications of the decision theory postulates. In K. Borch & J. Mossin (Eds.), Risk and Uncertainty. New York: St. Martin's, 1968.

MacCrimmon, K. R. & Larsson, S. Utility theory: Axioms vs. "paradoxes." In Rational Decisions Under Uncertainty, special volume of Theory and Decision, M. Allais & O.

Hagen, (Eds.), 1976.

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

Matsui, T. & Ikeda, H. Effectiveness of self-generated outcomes for improving predictions in expectancy theory research. Organizational Behavior and Human Performance, 1976, 17, 289-298.

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.

Moskowitz, H. Effects of problem representation and feedback on rational behavior in Allais and Morlat-type problems. Decision Science, 1974, 5, 225-242.

Muchinsky, P. M. The consistency of intrasubject valence and instrumentality measures: A methodological consideration. Academy of Management Journal, 1977, 20, 321-327.

Murphy, A., Lichtenstein, S., Fischhoff, B. & Winkler, R. L. Misinterpretations of precipitation probability forecasts. Bulletin of the American Meteorological Society, in press.

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

Peterson, C. R. & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.

Quinn, R. P. & Staines, G. L. The 1977 quality of employment survey. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1978.

Rosenberg, J. Cognitive structure and attitudinal affect. Journal of Abnormal and Social Psychology, 1956, 53, 367-372.

Salancik, G. & Pfeffer, J. An examination of need-satisfaction models of job attitudes. Administrative Science Quarterly, 1977, 22, 427-456.

Salancik, G. & Pfeffer, J. A social information processing approach to job attitudes and task design. Administrative Science Quarterly, 1978, 23, 224-253.

Schwab, D. P., Olian-Gottlieb, J. D. & Heneman, H. G. III. Between-subjects expectancy theory research: A statistical review of studies predicting effort and performance. Psychological Bulletin, 1979, 86, 139-147.

Simon, H. A. Models of Man. New York: Wiley, 1957.

Slovic, P. Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. Journal of Applied Psychology, 1969, 53, 255-263.

Slovic, P. Hypothesis testing in the learning of positive and negative linear functions. Organizational Behavior and Human Performance, 1974, 11, 368-376.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Cognitive processes and societal risk taking. In J. S. Carroll & J. W. Payne (Eds.), Cognition and social behavior. Potomac, Md.: Lawrence Erlbaum Associates, 1976.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Rating the risks. Environment, 1979, 21(4), 14-20, 36-39.

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, 458-479.

Slovic, P. & Tversky, A. Who accepts Savage's axiom? Behavioral Science, 1974, 10, 368-373.

Staw, B. M. Motivation from the bottom up. In B. M. Staw (Ed.), Psychological foundations of organizational behavior. Santa Monica, Ca.: Goodyear, 1977.

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

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

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

Wallsten, T. S. Subjective expected utility theory and subjects' probability estimates: Use of measurement-free techniques. Journal of Experimental Psychology, 1971, 88, 31-40.

White, S. E. & Mitchell, T. R. Job enrichment versus social cues: A comparison and competitive test. Journal of Applied Psychology, 1979, 64, 1-9.

Wilks, S. S. Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 1938, 3, 23-40.

Wright, G. N. & Phillips, L. D. Personality and probabilistic thinking. British Journal of Psychology, in press.

Wright, P. Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. Instructional Science, 1978, 7, 249-312.

Wyer, R. S. Cognitive organizationand change: An information processing approach. Potomac, MD.: Lawernce Erlbaum Associates, 1974.

Zagorski, M. A. Risky decision: Attention effects or masking effects? Acta Psychologica, 1975, 39, 487-494.

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER PTR-1091-80-4	2. GOVT ACCESSION NO. AD-A088 698	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The experienced utility of expected utility approaches		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Baruch Fischhoff, Bernard Goitein & Zur Shapira		8. CONTRACT OR GRANT NUMBER(s) [redacted] MDA-903-80-C-0194
9. PERFORMING ORGANIZATION NAME AND ADDRESS Decision Research A Branch of Perceptronics 1201 Oak Street, Eugene, Oregon 97401		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ARPA Order No. [redacted] 3831
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22217		12. REPORT DATE April 1980 13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

none

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Formal models
subjective expected utility

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Reviewing many research results, the present report concludes: (a) the ability of the weighted average models to predict the decisions that people make varies from situation to situation; (b) even where they are successful in predicting decisions, these models may not do a very good job of describing the psychological processes that determined them. That is, their predictive accuracy is in part an artifact of the research techniques used.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)