037230-2-T

THE UNIVERSITY OF MICHIGAN

AD 746156

ENGINEERING PSYCHOLOGY LABORATORY

Multi-Dimensional Value Assessment for Decision Making

Technical Report

GREGORY W. FISCHER

Prepared for:

Engineering Psychology Programs
Office of Naval Research
The Department of the Navy
Arlington, Virginia
Contract No. N00014-67-A-0181-0034
NR 197-014

Approved for Public Release; Distribution Unlimited

Administered through:

June 1972

[OFFI]CE OF RESEARCH ADMINISTRATION · ANN ARBOR

73

MULTI-DIMENSIONAL VALUE ASSESSMENT

FOR DECISION MAKING

Technical Report

1 June 1972

Gregory W. Fischer

Engineering Psychology Laboratory

The University of Michigan

Ann Arbor, Michigan

This research was supported by the Engineering Psychology Programs, Office of Naval Research, under Contract Number N00014-67-A-0181-0034, Work Unit Number NR 197-014.

1a

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Department of Psychology Univeristy of Michigan Ann Arbor, Michigan	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Multi-Dimensional Value Assessment for Decision Making

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Gregory W. Fischer

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 June 1972	68	89

8a. CONTRACT OR GRANT NO. N00014-67-A-0181-0034	9a. ORIGINATOR'S REPORT NUMBER(S) 037230-2-T
b. PROJECT NO. NR 197-014	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Engineering Psychology Programs Office of Naval Research

13. ABSTRACT

Decision analysis is a tool that can be used to improve the quality of complex decisions in an uncertain environment. A decision analysis is constructed by specifying alternative courses of action and the possible consequences of action. Each of the consequences is evaluated in terms of its relative probability of occurrence and its value to the decision maker if it should occur.

Decision analysis has been used primarily in business settings where values of consequences can be measured in terms of dollars. In non-business environments, however, non-monetary criteria may be of paramount importance. The situation is further complicated if relevant values vary along more than a single dimension. This paper reviews the psychological literature on the problem of assigning numerical values when several value attributes (or criteria) are relevant to the decision maker.

This literature is reviewed from both a descriptive and a normative point of view. That is, how do people evaluate multi-attribute objects, and how should they? A simple weighted average provides a good description of how people do, in fact, make such evaluations. The weighted average approach is also appropriate for many normative purposes and several procedures for making this evaluation process explicit are discussed and criticized.

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Multi-Attribute Utility						
Multi-Dimensional Utility						
Value Function						
Riskless Choice						
Decision Making						
Utility Assessment						
Evaluation Models						
Risk Attitude						

DD FORM 1 NOV 65 1473 (BACK)

ic

S/N 0101-807-5521

Security Classification

A-31409

TABLE OF CONTENTS

INTRODUCTION

Most significant decisions involve choosing between courses of action whose consequences involve multiple value relevant attributes. For example, in buying a car the value relevant factors might include price, appearance, steering and handling characteristics, fuel economy and resale price. Within the psychological literature the terms multi-attribute or multi-dimensional preference are used to refer to decisions of this type. This paper reviews the rapidly growing body of psychological research addressed to the following two questions: First, how do people assign value to multi-attribute outcomes? And second, what are the most useful procedures for obtaining a quantitative measure of the subjective worth of a multi-attribute outcome?

While most of this research is fairly abstract in nature, it is pragmatically oriented. The procedures which have been developed are designed for use in real world settings, and a number of studies have attempted to determine which of these procedures are likely to be most useful in real world contexts.

In organizing this literature it is useful to distinguish between risky and riskless decisions, between normative and descriptive theories of choice, and between intuitive and decomposed value judgments. Abstract theories of decision making assume that for each possible course of action there is one and only one outcome which will occur. At the time he makes his decision, however, the decision maker may or may not be aware of what that outcome will be. If the decision maker is able to specify with complete certainty the outcome associated with each course of action, then the decision is said to be riskless. A decision is said to be risky, on the other hand, if the decision maker is uncertain as to the consequences associated with each course of action but is able to express this uncertainty in the form of probability distributions over the possible consequences of each act.

The distinction between normative and descriptive theories of choice is somewhat arbitrary. Normative theories prescribe how the decision maker ought to make his choices. Typically, a set of basic principles of rational choice are postulated, then from these principles a rational strategy is deduced. Descriptive decision theory, on the other hand, is concerned with describing how people do in fact make decisions. But to the extent that human choice satisfies some or all of the principles embodied in a particular normative theory, then that theory will also be descriptive.

Finally, an evaluative process will be said to be intuitive if the synthesis of value relevant information is entirely subjective. Most decisions are made in this fashion. For example, after considering each of value relevant attributes of a job offer, the prospective employee will form an overall subjective impression of the desirability of the offer. In contrast, decomposed evaluation procedures are more explicit and rely upon mathematical rather than subjective synthesis of information. Broadly speaking these procedures involve four major steps. First, the decision maker must explicitly list the set of value relevant factors upon which he wishes to base his decision. Next, he must quantitatively assign relative importance weights to each of these factors. Third, the decision maker must numerically assess the value of each alternative outcome with respect to each of the value attributes. Finally, an arithmetic combination rule can be used to calculate the overall value of each alternative. In most cases a simple weighted average would be used.

The importance of the concepts described above is reflected in the organization of this paper. The two major sections deal with riskless and risky decisions respectively. Both sections begin with a discussion of descriptive studies of multi-attribute decision processes. Unfortunately, few studies of multi-attribute preference have been conducted in a risky choice setting. But for riskless choice a rather extensive body of research is available. Next, each section discusses normative approaches to multi-attribute evaluation problems. In both cases, special emphasis is placed upon the construction and validation of decomposed evaluation models.

MULTI-DIMENSIONAL VALUE ASSESSMENT IN THE ABSENCE OF RISK

Descriptive Theories of the Multi-dimensional Evaluative Process

A traditional economic treatment of riskless choice views the decision maker as being able to make continuous trade-offs between value attributes; these trade-offs produce the indifference maps upon which an economic analysis of the behavior of consumers and firms is based (Stigler, 1966). The implied psychological process is compensatory in the sense that an increase in value with respect to one attribute can compensate for a decrease in value on any other attribute.

Psychological studies of the multi-dimensional evaluative process have generally supported the compensatory trade-off model. These studies have typically utilized the following paradigm. First, the subject is asked to numerically assess the values of a set of multi-dimensional alternatives. Notationally, the i-th alternative can be described by the vector $(X_1^i, X_2^i, \ldots, X_n^i)$, where X_j^i denotes the j-th attribute of the i-th alternative. After the subject has evaluted all of the alternatives, the experimenter attempts to fit a statistical model to these judgments. A number of studies have shown that the human judgment process can be very well represented by additive compensatory models of the form $V(X_1^i, X_2^i, \ldots, X_n^i) = V_1(X_1^i) + V_2(X_2^i) + \ldots + V_n(X_n^i)$ where $V_j(X_j^i)$ is the value of the i-th alternative with respect to the j-th attribute.

Three experimental studies provide particularly strong support for this additive formulation. Tversky (1967) asked prison inmates to give minimum selling prices for commodity bundles consisting of m packs of cigarettes and n bags of candy. The hypothesis that the cigarettes and candy contributed additively to the overall value of the bundles was tested by analysis of variance. For none of the eleven subjects did the selling prices exhibit a statistically significant ($p \leq .10$) departure from additivity, and the mean within subject rank order correlation (Tau) between actual selling prices and those predicted by an additive model was .995. Sidowski and Anderson (1967) asked subjects to evaluate the attractiveness of job positions described by two attributes: type of work and city of employment. In two separate studies a statistically significant interaction was observed; but in both cases additive approximations accounted for almost all systematic variance ($R=.986$ and $R=.987$). Similar results were obtained by Shanteau and Anderson (1969) who asked their subjects to assess the value of various sandwich-soft drink combinations. Although significant interactions were obtained for five of twenty subjects, additive models gave a near perfect account of the data (mean $R=.998$).

Each of the studies discussed above utilized an analysis of variance design, with subjects evaluating all possible combinations of the stimulus attributes. This approach has two principle advantages. First, it provides a direct test of the additivity assumption (Anderson, 1970; Tversky, 1967). Second, the analysis of variance may be used as a scaling procedure to obtain the desired additive representation (Anderson, 1970). On the other hand, analysis of variance designs also have major drawbacks. Factorial combinations of attributes can yield attribute combinations which could never occur in the real world (Slovic and Lichtenstein, 1971). And, as the number of dimensions increases, the number of responses required of each subject rapidly becomes prohibitively large. It is noteworthy that all three studies above used only two value dimensions.

A number of investigators have attempted to bypass these difficulties through use of a multiple regression paradigm. Subjects respond only to a sample of the possible stimuli, then additive models are estimated by multiple regression. Studies in this tradition have adopted the rather restrictive assumption that each dimension contributes linearly to overall value; although non-linear regression methods are available, they require a large number of judgments by the subject. In addition, regression procedures provide no formal test of the additivity assumption, though the multiple correlation coefficient provides a convenient measure of goodness-of-fit. Finally, each independent variable must be represented as an interval scale number. Use of qualitative attributes is possible only if the experimenter arbitrarily assigns numerical values to these attributes.

Studies utilizing this paradigm have generally provided strong support for the additive model as a predictive tool. Huber, Sahney, and Ford (1969) asked hospital administrators to evaluate the quality of hospital wards described by seven attributes. All attributes were presented in numerical form; inherently qualitative factors, such as cleanliness, were assigned numerical scores ostensibly attained during the last hospital inspection.

After ratings of the wards were obtained the experimenters fit each of the following three models to this judgment data.

$$V_a = a_1 + a_2(X_1^i) + a_3(X_2^i) + \ldots + a_8(X_7^i)$$

$$V_b = b_1 + b_2\log(X_1^i) + b_3\log(X_2^i) + \ldots + b_8\log(X_7^i)$$

$$V_c = c_1(X_1^i)^{c_2}(X_2^i)^{c_3}\ldots(X_7^i)^{c_8}$$

Here X^i is the _i_-th ward, X_k^i is the _k_-th attribute of the _i_-th ward, and the a_k, b_k, and c_k are parameters to be estimated via least squares. All three models did an excellent job of accounting for the data, with median within subject correlations of .969, .973, and .976 respectively. This insensitivity to the algebraic form of the model is counterintuitive, but it is commonly obtained in studies of the judgment process.

In a similar study, Hoepfl and Huber (1970) asked engineering graduate students and faculty to evaluate the teaching ability of hypothetical professors. These judgments were based on from one to six attributes. Again linear regression models did an excellent job of explaining the data. In general, however, the degree of correlation declined (from .987 to .940) as the number of attributes increased from one to six. Hoepfl and Huber noted that this decline might have been due to some sort of information overload effect. As the amount of information to be processed by the subject increased, his responses became increasingly subject to random error. Huber, Daneshgar, and Ford (1971) obtained less favorable results in a study of the job preferences of prospective public school teachers. The study used a mailed questionnaire as a response device, and subjects gave ratings for jobs described by five attributes. Mean correlations of .89 and .71 were obtained for subjects with and without prior teaching experience respectively.

In summary, the studies above indicate that simple additive models can do a remarkably good job of approximating the human evaluative process. With the exception of the Huber, Daneshgar, and Ford (1971) study, median within subject correlations for best-fitting additive models were in the mid to high .90s. As Anderson (1969) has noted, however, these high correlations do not necessarily imply the absence of any significant interactions. In fact, two of the three studies which provided a formal test of the additivity assumption found evidence of interactions. Even in these cases, however, additive models exhausted almost all of the predictable variance.

Moreover, additive models have been found to be highly descriptive of a variety of other human judgment processes. Personality impression studies are concerned with the process whereby people synthesize information about specific characteristics of another person in forming an overall impression of that person. In general, these studies have found that overall judgments of the attractiveness of a person can be predicted by additive combinations of his specific characteristics (Anderson, 1962; Anderson and Jacobson, 1965; Anderson, 1967; Himmelfarb and Senn, 1969). The additive model has also received support from studies of clinical judgment (Meehl, 1954; Golberg, 1968; Golberg, 1970), investment decision making (Slovic, 1969), graduate admissions decisions (Dawes, 1970), governmental budgetary decisions (Dempster, Davis, and Wildavsky, 1971; Crecine and Fischer, 1971), and corporate decision making (Bowman, 1963). This literature has been critically reviewed by Slovic and Lichtenstein (1971).

Despite the impressive predictive power of additive compensatory models, which permit cross-dimensional trade-offs, a number of psychologists have questioned whether they are truly descriptive of the processes underlying multi-dimensional choice. Skeptics generally argue that decision makers utilize non-compensatory heuristic evaluation rules. In addition, it is argued that people employ different heuristics depending upon the context in which the decision is made.

Students of organizational behavior have argued that decision makers frequently utilize a satisficing (or conjunctive) strategy (March and Simon, 1958; Cyert and March, 1963). In employing this strategy, the decision maker (DM) first establishes a minimum acceptable level with respect to each value attribute. Any course of action which fails to satisfy one or more of these minimal constraints is rejected as unacceptable. So given a set of alternative courses of action, a satisficing rule will partition it into two subsets: those which are acceptable and those which are not. If two or more alternatives are acceptable, additional considerations must be introduced in order to choose between them. For example, DM might choose that acceptable alternative which is "best" with regard to the most important attribute. If no alternative passes the test of admissibility, DM must either search for new alternatives or relax one or more of his criteria. A disjunctive strategy, on the other hand, is concerned not with acceptability, but rather with excellence. Each alternative is evaluated only with respect to its most outstanding attribute, and that alternative whose best attribute is most desirable is selected. For example, if a disjunctive strategy were used by a college admissions office, a student with poor mathematical ability but excellent verbal ability would be favored over a student who was average in both regards.

Einhorn (1970,1971) has attempted to contrast the predictive power of these heuristic strategies with that of additive compensatory models. In his work Einhorn has used the following multiplicative approximations to the conjunctive and disjunctive strategies respectively.

$$V_c = (X_1^i)^{a_1}(X_2^i)^{a_2}\ldots(X_n^i)^{a_n}$$

$$V_d = \left[1/(c_1-X_1^i)\right]^{b_1}\left[1/(c_2-X_2^i)\right]^{b_2}\ldots\left[1/(c_n-X_n^i)\right]^{b_n}$$

Here the $\underline{c}_k$ are constants such that $\underline{c}_k > \max(X_k)$, and the $\underline{a}_k$ and $\underline{b}_k$ are weighting parameters to be statistically estimated. The rationale for these models is that they reflect the qualitative properties of the conjunctive and disjunctive strategies. But in contrast to the strategies which they represent, the equations are compensatory in nature. Moreover, the equations are additive under logarithmic transformations. Thus, Einhorn has, in effect, compared linear additive models with non-linear additive models. In addition, he has demonstrated that additive models are capable of reflecting a wide range of qualitative properties.

Einhorn employed a double cross validated paradigm in his experimental studies. Subjects responded to two stimulus sets, with models estimated on one set used to predict responses to the other set. In his first study, dealing with job preferences, Einhorn found that the "conjunctive" model gave the best predictions. In the second study, involving faculty evaluations of prospective graduate students, a simple linear additive model tended to do as well as or better than the two non-linear versions. In view of the fact that the double cross validated design involved real prediction of new data, the obtained correlations (Rho) of .85 and .81 were quite respectable, and provide additional evidence of the representational power of additive models. Einhorn's procedures do not, however, allow one to determine whether or not subjects were utilizing non-compensatory strategies.

Lexicographic rules constitute a third class of non-compensatory heuristic strategies. To employ this approach DM first compares all alternatives with respect to the most important value attribute. If one alternative dominates all others with respect to this criterion, it is selected. If two or more alternatives are equivalent at this stage, they are compared with respect to the next most important value attribute, and so on, until only one alternative remains. Although this strategy is very inefficient in the sense of systematically ignoring value relevant attributes of alternatives, Tversky (1969) has found that some subjects consistently employ such a strategy. In contrast to other studies in which subjects were asked to rate alternatives, subjects in these studies were asked to choose between pairs of alternatives.

A fourth heuristic model, developed by Tversky (1971) and termed the Elimination By Aspects (EBA) model, shares certain features of both the conjunctive and lexicographic rules. According to the EBA model each alternative consists of a set of binary aspects or attributes. At each stage of the decision process an attribute is selected with probability proportional to its importance. All alternatives which are unsatisfactory with respect to this aspect are eliminated from further consideration. (Of course, if all are unacceptable, none are eliminated.) Additional attributes are probabilistically selected until all but one alternative is eliminated. Tversky (1971) has shown that EBA overcomes major shortcomings of other probabilistic choice theories. This theoretical superiority stems from the fact that attributes shared by all alternatives do not effect the choice probabilities. Tversky has conducted studies testing the implications of the EBA model and has obtained fairly strong support for this formulation. Here again, subjects were asked to choose between alternatives rather than to assign ratings to them.

Tversky's evidence of non-compensatory decision making is difficult to reconcile with the rather large number of studies for which additive compensatory models have done so well. One possibility is that people use compensatory strategies when they bid or rate, but use non-compensatory heuristics when they choose or rank-order. It is also possible that the models in question are "paramorphically equivalent". Hoffman (1960) has noted that it may be the case that algebraically different models, each suggestive of different underlying processes, may be equally predictive given fallible data. At present the issue is unresolved.

Aiding the Decision Maker: Normative Procedures for Multi-dimensional Value Assessment.

This paper uses the word "normative" in a fairly loose sense. The procedures to be discussed do not attempt to produce decisions which are optimal in some absolute sense. Their goal is, rather, to enable decision makers to make better choices.

Simon (1969), Dawes (1964), and MacKrimmon (1968) have argued the normative merits of heuristic strategies. But, as both Raiffa (1969) and Tversky (1971) have argued, the non-compensatory nature of these procedures seems extremely undesirable from a normative standpoint. Thus, we consider only compensatory approaches.

Bootstrapping procedures are an offshoot of attempts to apply linear statistical models for descriptive purposes. The essence of this approach is to replace the decision maker with a model of the decision maker. Decomposition strategies, on the other hand, require the decision maker to make value judgments about individual value attributes and then combine these judgments according to some arithmetic combination rule. Both procedures are intended not only to relieve the decision maker of some of his burdens but also to produce "better" decisions.

The Bootstrapping Technique. We found earlier that linear statistical models could do a good job of reproducing a decision maker's evaluations of multi-attributed outcomes. Dawes (1970), Goldberg (1970), Bowman (1963), and others have argued that it might be useful to replace the decision maker with a model inferred from his previous judgments or decisions. In routine decision making contexts this would free the decision maker to apply his intellectual abilities to more challenging tasks (Yntema and Torgerson, 1961). Once developed, a computerized evaluation model could rapidly calculate the values of thousands of alternatives. In addition, it has been argued that bootstrapping can lead to improved decisions. Underlying this assertion is the assumption that random error is a major source of non-optimality in human judgment (Bowman, 1963). Subjective weighting of dimensions, for example, might be overly responsive to momentary environmental events, or decision makers may be erratic in aggregating information across attributes. Statistical models, however, can extract systematic effects while filtering out noise due to random error.

The assertion that bootstrapping models can improve upon the quality of the decision process has received some support from real world studies. Goldberg (1968, 1970) has found that bootstrapping can be used to improve the quality of medical diagnoses. Similar results have been obtained in the context of business decision making (Bowman, 1963) and the selection of graduate students (Dawes, 1971).

Although the full potential of bootstrapping has yet to be determined, certain limitations seem inherent in the procedure. First, it is ill-suited for application to complex decisions without precedent. For unless the decision maker is able to assign overall values to a fairly large set of hypothetical outcomes, models cannot be estimated. And when faced with complex multi-faceted outcomes decision makers may well be unsure of their preferences and thus unable to make the required assessments.

Bootstrapping is also ill-suited to decisions involving a large number of value relevant dimensions. For a number of studies have shown that intuitive judgments (upon which the bootstrapping model must be based) tend to focus very heavily upon but a few dimensions of value (Slovic and Lichtenstein, 1971), so that use of a bootstrapping procedure for a problem involving ten or more attributes would essentially result in throwing away value relevant information. Thus, application of bootstrapping will probably be confined to routine decision making contexts involving many decisions but few dimensions of value.

The Decomposition Approach. Bootstrapping procedures rest on the assumption that basically the decision maker knows what he is doing, but that he makes his judgments in a noisy fashion. Decomposition, on the other hand, is based upon the assumption that decision makers are not well equipped for the task of evaluating complex multi-dimensional outcomes. Shepard was one of the first proponents of this approach. He argued that while human sensory and motor capacities were developed to a high degree, man's ability to process conceptual information was much less impressive.

Studies of human information processing, for example, indicate that people can process only five to ten "chunks" of information at any one time (Miller, 1956; Norman, 1969; Fitts and Posner, 1969). People also display little capacity to learn concepts based on three or more interacting attributes (Shepard, Hovland, and Jenkins, 1961). The limits on the human capacity to make multi-dimensional judgments are perhaps best illustrated by the literature on human clinical judgment. Medical diagnosis typically requires physicians to categorize patients on the basis of a set of signs, symptoms, and test results. When asked how they make such judgments, physicians report that they take into account complex interrelations between the indicators available to them. Yet statistical analyses of the clinical judgment process have consistently found that linear regression models account for almost all of the consistent variance in these judgments (Goldberg, 1968; Slovic and Lichtenstein, 1971). Moreover, human clinical judgments are not particularly accurate. Meehl (1954) found that linear statistical models outperformed trained diagnosticians. Slovic, Rorer and Hoffman (1971) found low agreement between radiologists assessing the same cases. Worse, even extensive training with feedback on thousands of case studies did little to improve the accuracy of clinical judgments (Goldberg, 1968). For psychologists working in this tradition it has seemed natural to assume that decision makers are also poorly equipped to make value judgments across multiple criteria. Decomposition procedures have been devised to improve the quality of the multi-dimensional evaluative process.

The following discussion considers first two methods for obtaining a decomposed value measure, and next the problem of validating such a measure. The two scaling procedures discussed are fairly representative of those generally proposed, and both assume that values are additive. The first is based upon direct rating scale judgments; the second upon trade-off or indifference judgments. The "additive rating scale" procedure is adapted from Edwards (1971), Fishburn (1965), Sayeki (1970), and Hoepfl and Huber (1970), and entails four major steps.

1. Within each dimension, the decision maker (DM) must specify the best and worst outcomes which could feasibly arise. When the set of alternatives is specified prior to the analysis, this is trivial. But it decisions are to be made over an extended period of time and applied to as yet unspecified options, then good judgment is required in selecting these endpoints. Within each dimension arbitrary values of 100 and 0 may be assigned to the best and worst feasible outcomes respectively.

2. Within each dimension, DM must assign numerical values to all outcomes intermediate in value to the best and worst. These numerical assessments should accurately reflect value differences within a given dimension; but they need not reflect value differences across dimensions. (In general, they will not.) A large number of procedures have been devised for obtaining interval scale utility judgments within a single dimension of worth (Fishburn, 1967), but in practice direct rating procedures have generally been used. For each possible intermediate outcome on a dimension, the DM is asked to

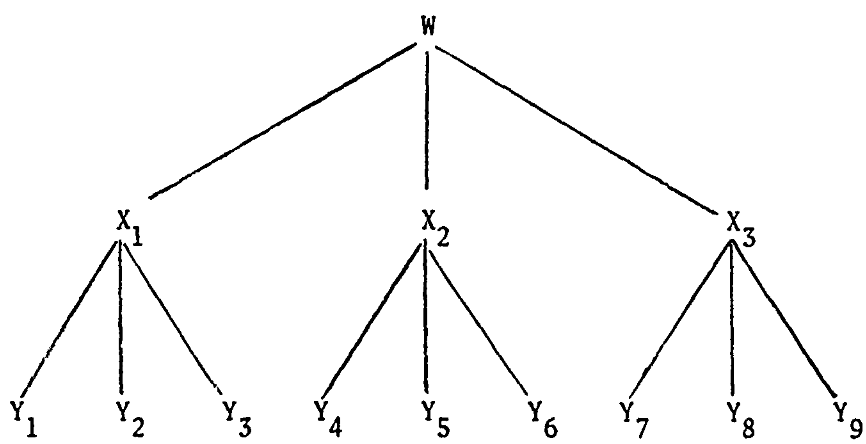
assess a number between 0 and 100 which reflects the subjective value of the outcome in question relative to the worst and best outcomes on that dimension. When the value dimension is continuous, interpolation between well chosen points is necessary.

3. Next, DM must assign weighting factors reflecting the relative contribution of each dimension to overall value. The weight assigned to an attribute should reflect the range of value produced by moving that attribute from its lower bound to its upper bound. For example, in choosing a job, the weight attached to salary might be relatively small if all offers ranged in salary from $10000 to $11000; but, all other things being equal, if salary ranged from $10000 to $15000, this factor might assume a considerably greater degree of importance. Quantitative assessments of these weights can be obtained by having DM assign a weight of 100 to the attribute with the greatest value range. All other attributes are then assigned weights in proportion to their associated value ranges. For convenience these weights can then be normalized to sum to one.

4. A value may now be assigned to any multi-attribute outcome by aggregating the weights and values obtained above according to the additive combination rule $V(X_1^i, X_2^i, \ldots, X_n^i) = \sum_j w_j V(X_j^i)$. If a decision is to be made, that action is chosen whose associated outcome has the greatest value.

A variant of the additive rating scale approach treats values as being organized in a hierarchy of goals, subgoals, sub-subgoals, and so on.

Miller (1969) and Sayeki and Vesper (1971) have developed procedures for constructing additive models incorporating a hierarchical value structure. In practice these procedures differ from the additive rating scale method primarily in terms of the manner in which weights are assessed. Consider the simple value hierarchy displayed below. W is the overall goal, and X_1, X_2, and X_3 the major goals leading to W. Y_1, Y_2, and Y_3 are subgoals leading to X_1; Y_4, Y_5, and Y_6 are subgoals leading to X_2; and so on.



First, importance is allocated across the major goals X_1, X_2, and X_3 and normalized to sum to one. Suppose for the purpose of illustration that weights of .5, .3, and .2 are allocated to X_1, X_2, and X_3 respectively. Next, within each goal, importance is allocated across subgoals. For example, suppose that within goal X_1, relative weights of .6, .3, and .1 are assigned to

Y_1, Y_2, and Y_3 respectively. Final weighting factors for the measurable attributes at the bottom of the hierarchy may be obtained by multiplying weights downward through the hierarchy. In the example, attribute Y_1 would receive a final weighting of $(.5)\cdot(.6) = .3$; attribute Y_2 a final weighting of $(.5)\cdot(.3) = .15$; and so on. In principle, weights obtained through a hierarchical procedure should be the same as those obtained by direct assessment of the attributes at the bottom of the hierarchy. In practice this may not be the case. The hierarchical approach is most attractive in situations where the "real" goals are vague and ill-defined. The hierarchy may then be used to select as attributes operationally specifiable subgoals which are indicators of the decision maker's real objectives.

The second decomposition procedure is adapted from Raiffa (1968, 1969) and Toda (1971) and will be termed the "additive trade-off method". In order to reduce notational complexity we shall consider only the three dimensional case, but the procedure can easily be extended to a larger number of dimensions. Let each outcome be characterized by the attribute vector (a^i, b^j, c^k), and assume that the first attribute, <u>a</u>, is continuous and can reasonably assume a wide range of values. This attribute will serve as a base against which the other factors will be traded off. In making these trade-offs one can exploit a property of additive models alternatively referred to as <u>monotonicity</u> (Yntema and Torgerson, 1961), <u>independence</u> (Adams and Fagot, 1959), or <u>weak conditional utility</u>

independence (Raiffa, 1969). The basic idea is that preferences for outcomes on any attribute or subset of attributes be unaffected by the state of other attributes. A formal definition will be provided later, but a simple example will illustrate this idea. Consider outcomes on the a attribute. If for any b', c', $(a^i,b',c') \precsim (a^j,b',c')$, then for all b", c", $(a^i,b'',c'') \precsim (a^j,b'',c'')$, where $X \precsim Y$ is interpreted as "X is not preferred to Y". This property assures that trade-offs between any two dimensions do not depend upon the state assumed by other dimensions (Raiffa, 1969). The additive trade-off decomposition relies heavily upon this property and may be decomposed into five major steps.

1. DM begins by trading off the b and c dimensions against the base dimension, a. Arbitrarily we establish "standard outcomes" for each of the three dimensions denoted by a^o, b^o, and c^o. These standard outcomes have no particular significance in theory, but in practice should be selected so as to keep all judgments within the range of outcomes which might reasonably occur. In trading off b against a we assume that a is continuous and that for each outcome b^i DM can specify an a^i such that $(a^o,b^i,c^o) \sim (a^i,b^o,c^o)$, where $X \sim Y$ is to be interpreted as "DM is indifferent between X and Y". Intuitively, a^i should be selected such that the difference in value between a^i and a^o is equal to that between b^i and b^o. Thus, value differences along the b dimension can be expressed in terms of units of a. Similarly, c can be traded off against a so that for each c^j and a^j is specified such that $(a^o,b^o,c^j) \sim (a^j,b^o,c^o)$.

2. Next, we arbitrarily establish the standard outcome, (a^o,b^o,c^o), as the origin of our desired value measure, so that $V(a^o,b^o,c^o) = 0$. All other outcomes will be compared with this standard outcome. If the standard outcome is preferred, then the value of the outcome in question will be negative; if the outcome in question is preferred, then its value will be positive.

3. DM must next assess a value function over the base dimension, a, with b and c at their standard levels so that a^o must be the zero point of this function over a. Thus, the value function over a may assume negative values. This function can be obtained by any of the scaling methods described by Fishburn (1967); again, a direct rating method should be satisfactory.

4. Outcomes on the b and c dimensions may now be scaled in terms of the value measure on a. For any b^j, for example, the trade-offs of the first step specify an a^j such that (a^o,b^j,c^o) is equivalent in value to (a^j,b^o,c^o). Thus, by the assumption of additive values

$$V(a^o) + V(b^j) + V(c^o) = V(a^j) + V(b^o) + V(c^o).$$

So, by the prior specification of the origin, $V(b^j) = V(a^j)$. Similarly, for each c^k there is an a^k such that $(a^o,b^o,c^k) \sim (a^k,b^o,c^o)$ so that $V(c^k) = V(a^k)$.

5. Combining these results, the value of any outcome (a^i,b^j,c^k) can be computed from $V(a^i,b^j,c^k) = V(a^i) + V(a^j) + V(a^k)$, where a^j and a^k are obtained from the trade-off stage. Again, if a decision is to be made, DM

should choose that course of action whose associated outcome has the greatest value.

The additive trade-off decomposition is especially suited to problems involving an important continuous dimension such as dollars or lives saved. Logically, it is equivalent to the rating scale procedure described earlier, and if properly carried out the two methods should produce identical decisions. Extensions of trade-off techniques to the non-additive case have been discussed by Toda (1971).

Validating a Decomposed Value Measure. The procedures described above are not difficult to carry out. They have, in fact, been occasionally employed in real world decision making contexts. Nevertheless, an important question remains. Do these procedures provide an "appropriate" measure of value? This problem of validation has been attacked from several rather divergent perspectives.

When an objective (externally defined) measure of value is available, validation is straightforward. For example, in certain medical diagnosis contexts, bootstrapping techniques have been shown to do as well as or better than physicians. But where value is concerned, such external criteria are typically unavailable. Yntema and Torgerson (1961) argued, however, that value measurement techniques can be experimentally evaluated by creating a simulated decision making environment governed by an arbitrarily specified value structure. They trained subjects to evaluate the "worth" of geometrical stimuli varying on three dimensions. Arbitrarily, the experimenters established a non-additive Worth function assigning a score to each stimulus on the basis of its three attributes. Each dimension was monotonically related to overall Worth, and an additive main effects approximation to the true rule accounted for 94% of the variance. On each training trial subjects estimated the Worth of a stimulus and were then given quantitative feedback on its true Worth. After extensive training, test trials were conducted on which subjects received no feedback. On these test trials, the mean correlation between subjects' estimates and true Worth was .84. Finally, for each subject a decomposed value model was constructed and used to predict the Worths of the test stimuli; here a correlation of .89 with true Worth was obtained. A bootstrapping model, based on subjects' responses to the test stimuli, also outperformed the intuitive judgments of the subjects; again a correlation of .89 was obtained. These results provide mild encouragement to those who believe that decomposition or bootstrapping can improve the quality of decision making.

Stronger support for the decomposition approach is provided by a recent study conducted by Lathrop and Peters (1969). The experimenters had previously collected course evaluation sheets from fourteen introductory psychology classes. On these sheets students had evaluated the course with respect to each of fifteen attributes. In addition, students had assigned an overall rating to both the course and the instructor. These actual student ratings were averaged, and the average scores treated as an objective value measure. Subjects in the experiments were not members of these classes, and their task was to predict the evaluative ratings of the students who had been. Two response modes were used. In the intuitive mode, subjects were presented with average class evaluations with respect to each of the fifteen factors and asked to predict the average overall evaluations. In the decomposed condition subjects assigned weights to each of the fifteen attributes. These weights were to reflect the subjects' best estimate of the relative importance which an average student would assign to each of the fifteen factors. These weights were combined with the actual attribute ratings to obtain a decomposed linear additive prediction of the actual overall ratings. For both teachers and courses, decomposed models afforded better prediction than did the intuitive judgments. For instructor evaluations, correlations of .973 and .896 were obtained for the decomposed and intuitive predictions respectively; for class evaluations, corresponding correlations of .963 and .924 were obtained. Lathrop and Peters also found, however, that the weights assessed in the decomposed mode were decidedly non-optimal when compared with weights derived from multiple regression. The decomposed weights were relatively uniformly distributed across the attributes, whereas the optimal statistical weights were much more heavily concentrated on but a few factors. In this experiment, the superiority of decomposition seems primarily to have arisen because decomposed judgments are less noisy than intuitive judgments aggregated across fifteen dimensions, a result consistent with the information overload hypothesis. Although it may seem non-intuitive that a decomposed model with badly estimated weights can yield excellent predictions, O'Connor (1972) has found that additive models are amazingly insensitive to changes in weights. Nevertheless, practical implementation of decomposition clearly requires additional consideration of the weight estimation problem.

Other approaches to the validation problem have also been proposed. Miller, Kaplan, and Edwards (1967, 1969), working in a military context, have argued that reliability over time is a minimal requisite of any value measurement procedure.

> If a subject's value judgments collected at any one time systematically differ from his value judgments for the same target in the same situation collected at a different time, there would be some doubt about the appropriateness of implementing either set of values (1967, p. 364).

Miller, Kaplan, and Edwards also discuss the usefulness of convergence (or "construct validity") as a validating principle.

> The basic idea of construct validity is that a test should make sense and the data obtained by means of it should make sense. One form of making sense is that different procedures purporting to measure the same abstract quantity should covary (1967, p.367).

Convergence has, in fact, been the criterion most commonly employed by psychologists. Bootstrapping models, for example, may be validated by noting that they reproduce the systematic components of a decision maker's intuitive value judgments. Prediction of intuitive judgments has also been used to validate decomposition procedures. This may seem to contradict the usual assertion that decomposed models are superior to intuition. But given the relative insensitivity to weights in the additive model, it seems reasonable to expect a high correlation between intuitive and decomposed ratings, at least for a small number of dimensions. In addition, it seems reasonable to demand a high degree of convergence between logically equivalent decomposition procedures.

Pollack (1964) was the first to employ this procedure. First, subjects rated and ranked jobs described by eight attributes, then additive decomposed value models were developed. Rather surprisingly, Pollack found that decomposed models incorporating all eight attributes did a poorer job of predicting the intuitive ratings (mean R=.7) than did a decimposed model including only the three most important attributes (mean R=.8). A closer examination of the data revealed that intuitive preferences were based only upon the three most important factors, whereas the decomposed models assigned some importance to all factors. Yntema and Klem (1965) conducted a similar study in which experienced pilots evaluated landing situations described by three attributes. Here, decomposed models did a good job of predicting intuitive preferences.

Hoepfl and Huber (1970) asked engineering graduate students and faculty to evaluate the teaching ability of hypothetical professors described by from one to six attributes. In the first stage of the experiment, intuitive judgments were obtained. Next, each subject constructed an additive decomposed value model using the rating scale procedure. These decomposed models were then used to predict the previously collected intuitive judgments. Median correlations between decomposed and intuitive judgments ranged from .87 to .98, with the degree of correlation generally declining as the number of attributes increased. Again, the decomposed weights differed significantly from the optimal regression weights which were concentrated more heavily on a few key dimensions.

Huber, Daneshgar, and Ford (1971) asked prospective school teachers to give intuitive judgments of the attractiveness of hypothetical jobs, then developed rating scale decomposition models. Both tasks were accomplished by mailed questionnaires. In this study the convergence between intuitive and decomposed ratings was much lower, with median correlations of .62 and .67 for those with and without prior teaching experience respectively. Although these results are somewhat discouraging, it should be noted that this was also the only study for which statistical models of the subjects failed to give a good account of the intuitive judgments. Apparently, the experimental procedures used introduced a large amount of error variance into the intuitive judgments.

Pai, Gustafson, and Kiner (1971) evaluated the predictive power of three decomposition procedures. First, subjects rank ordered the attractiveness of ten universities described by four attributes. Next, three decomposition models were constructed. In the first procedure, scales within dimensions were obtained by having subjects draw them on a sheet of graph paper. The second method obtained scales within dimensions by means of ratio type direct estimates. The final procedure required a set of fairly complex cross dimensional judgments and was designed to cope with possible non-additivity. A correlational analysis revealed that the "draw a curve" and ratio based additive decompositions were approximately equally predictive of the intuitive rank orderings (mean rank order correlations of .81 and .77), but that the non-additive method was substantially poorer.

Von Winterfeldt (1971) also compared the predictive power of two decomposition procedures. First, subjects ranked a set of apartments described by fifteen attributes. Then two additive decomposition models were developed. The first model used weights obtained in the manner described for the rating scale method. The second model used weights obtained by direct assessments of the value ranges associated with each attribute. Then, after extensive discussions of the relevant aspects of the apartments, subjects reranked the set of apartments. For both decomposition methods, median correlations with the second set of rankings were in the low .80's. In view of the number of dimensions involved, the degree of convergence obtained is quite impressive. Possibly the procedures encouraged subjects to consider more factors in making their intuitive judgments than would usually be the case.

Fischer (in prep.) has conducted two studies, both of which yielded a high degree of convergence between decomposed and intuitive judgments. Subjects in the first study evaluated hypothetical compact cars described by either three or nine attributes. Each subject utilized two intuitive response modes. In the Intuitive Rating Scale mode subjects rated a set of compact cars on a 0 to 100 scale. For the Intuitive Dollar Difference mode, subjects compared each car in the stimulus set with a "standard car" which was approximately average in all respects. After specifying which of these two cars he preferred, the subject assessed the dollar difference in value between the two cars. Each subject also constructed two decomposed value models. The first of these was an Additive Rating Scale, decomposition of the type described earlier. Subjects also constructed an Additive Dollar Trade-off model. Here, a slightly modified version of the previously discussed additive trade-off method was used. Within dimension value differences were traded off into dollars. These dollar differences were then assumed to combine additively across dimensions.

All intuitive judgments were obtained before the decomposed models were developed. These models were then used to predict the intuitive judgments. Intuitive Rating Scale judgments were correlated with the values predicted by the Rating Scale decomposition; Intuitive Dollar Difference judgments with the values predicted by the Dollar Trade-off decomposition. For the Rating Scale response mode, median within subject correlations of .95 and .93 were obtained for three and nine dimensions respectively. For the Dollar Difference response mode, corresponding correlations of .92 and .97 were obtained. These results suggest that, to a fairly high degree, intuitive and decomposed value measures are tapping the same underlying attribute - subjective value.

This experimental design also permitted a second application of the convergent validity criterion. As was noted earlier, the two decomposition procedures are in principle equivalent, and should lead to the same decisions. The same ought to be true of the two intuitive value measures. Below are the results of a correlational analysis which contrasted the degree of convergence between the two decomposed measures with that between the two intuitive measures.

	Three Dimensions	Nine Dimensions
Median Correlation between Intuitive Measures	.92	.95
Median Correlation between Decomposed Measures	.98	.97

Although all of these values are encouragingly high, decomposed measures showed a higher degree of convergence for both three and nine dimensions.

In the second experiment, subjects evaluated job offers described by three attributes--city, salary, and type of work. Because the experiment was primarily concerned with risky choice, only one riskless response mode was used, the 0 to 100 scale. Each attribute assumed only three states, and subjects evaluated all of 27 possible combinations of these attributes. Next, an additive rating scale decomposition was constructed. Again, the intuitive and decomposed methods yielded strikingly similar results (median R=.965). All subjects were experienced professionals, and all decision tasks were in the subjects' fields of expertise. Subjects assigned weights to six criteria using six different weight estimation procedures. One procedure relied solely on rank orderings, one on ratings, three on paired comparisions, and the last was an iterative procedure requiring successive comparisons and ratings. Within judge reliability across weighting procedures was very high. For engineering design tasks, correlations were typically between .99 and 1.0; for personnel selection tasks, correlations were only slightly lower.

Sayeki and Vesper (1971), on the other hand, found that hierarchical procedures generated systematically different weights than those obtained by direct evaluation of the attributes at the bottom of the hierarchy. The degree of discrepancy increased as the number of levels in the hierarchy increased.

Huber, Daneshgar, and Ford (1971) suggested a third and final criterion for validating multi-dimensional riskless value measures. In their study of the job preferences of prospective public school teachers, they developed both bootstrapping and decomposed value models. Subsequently, the authors obtained data on the set of job offers actually received by each teacher, and on the job selected from that set by the teacher. Thus, it was possible to evaluate both the bootstrapping and decomposition models in terms of their ability to predict real decisions. For the experienced teachers, decomposition models predicted the job selected in ten of fifteen cases; bootstrapping models predicted only seven of fifteen choices. Results were generally poorer for those job candidates without prior teaching experience, but again the decomposition models were better, predicting eight of fifteen choices as contrasted with only two of fifteen for the bootstrapping models. Huber, Daneshgar, and Ford also computed the proportion of cases for which the job actually selected fell in the upper quartile of the ratings produced by the evaluation models. For the experienced teachers, twelve of fifteen fell in the upper quartile of the decomposed ratings, only ten in the upper quartile of the bootstrapping models. For inexperienced teachers, comparable figures of ten and seven of fifteen were obtained. In general, the decomposed evaluation models did a good job of predicting the real job choices of the teachers, especially for those with prior experience. This predictive power is all the more outstanding in view of the fact that the real choices may have been influenced by factors not included in the decomposed models. For example, some teachers had to take jobs in the towns where their husbands resided.

To summarize, three approaches to the validation problem have been considered. Although the data are sparse, comparisons with external value measures or with subsequent decisions of major importance have revealed a high degree of convergence. Comparisons between intuitive and decomposed judgments, for which a fair degree of data is now available, have produced somewhat more mixed results. In general, correlations have been in the .80s or .90s, but poorer results were obtained by Pollack (1964) and Huber, Daneshgar, and Ford (1971). In the latter case, however, there was evidence that the poor convergence may have been attributable to noise inherent in the mailed questionnaire response device. Future studies, particularly those in real world contexts, should incorporate reliability measures on the intuitive judgments which are to be used as validating criteria. Comparisons of different decomposition methods have revealed a high degree of convergence between non-hierarchical models, but there is some evidence of poor convergence between hierarchical and non-hierarchical models. If the latter result is confirmed, additional studies utilizing external validating criteria will be required to determine the relative merits of hierarchical and non-hierarchical procedures.

Despite these generally encouraging results, however, one important problem remains unresolved. Decomposition studies have consistently found that the weights obtained via decomposition tend to be fairly flatly distributed across attributes. Decomposed weights are flatly distributed not only relative to intuitive weights (Pollack, 1964; Hoepfl, and Huber, 1970) but also relative to optimal statistical weights in the presence of a known criterion (Lathrop and Peters, 1969). Similar problems have been encountered in O'Connor's (1972) attempt to apply decomposition procedures for the development of measures of water quality. Water pollution experts scaled functions over important water parameters and then assigned weights to each of these factors. The initially estimated weights gave a ratio between weights of the most and least important factors of only 1.7 to 1. In view of the fact that the most important factor (fecal coliforms) involved a potentially severe health hazard while the least important factor (color) was of only minor aesthetic significance, this ratio was viewed as unreasonably small. After discussing this problem with his pollution experts, O'Connor was able to obtain a greater degree of variation in the magnitudes of the weighting factors.

But despite the apparent non-optimality of the weights estimated, decomposition models have generally done a good job of predicting validating criteria. O'Connor conducted a sensitivity analysis for his problem and found that additive models are remarkably insensitive to minor (roughly monotone) variations in weights.

While further analysis is required, these results cast a new light on decomposition. They suggest that the principal advantage of decomposition over intuition is that it eliminates error variability from the judgment process, and that any mechanically implemented model may do a fairly good job. Thus, for small numbers of dimensions there is little reason to expect major differences in the relative quality of bootstrapping and decomposition. For a large number of attributes an interesting question is raised. Bootstrapping is viewed as unsuitable here because it will assign essentially zero weight to all but a few attributes. But decomposition may err in the direction of assigning too much importance to trivial factors. This problem can only be studied in contexts where an external validating device is available. It also seems that decomposed weighting procedures should be evaluated in terms of their ability to produce weights whose flatness is sensitive to the relative distribution of weights for the externally specified data generator. Finally, additional numerical analyses are called for to determine how serious the weighting problem is.

Practical advantages of decomposition. We noted earlier that bootstrapping might be useful in decision making contexts requiring routine evaluation of the worth of many outcomes. Thus conceived, bootstrapping is primarily a labor saving device designed to free decision makers for more interesting tasks. In addition, it was argued (and shown) that bootstrapping can improve the quality of the evaluative process by filtering out random error. Decomposition procedures share these virtues. In addition, decomposition has several advantages over the bootstrapping approach.

First, decomposition will typically require fewer and easier judgments on the part of the decision maker. Using regression procedures for bootstrapping the stability of weighting coefficients depends upon the number of judgments upon which the model is based. And the number of judgments required increases with the number of coefficients to be estimated. For an evaluation problem involving ten attributes, thirty judgments seems an absolutely minimal basis for model estimation. Laboratory experience suggests that decision makers will find this a very difficult and time consuming task at best, and an impossible task at worst. In contrast, a rating scale or trade-off decomposition over ten attributes can be developed with relative ease.

Second, decomposition can be applied in contexts where the number of attributes is so large as to render bootstrapping useless.

Third, decomposition can be applied in contexts where the decision maker is unsure of his preferences. In many cases a decision maker will be hard pressed to consider the merits of two or three multi-attributed outcomes, much less to assign interval scale evaluations to fifty such outcomes. This is especially likely to be the case when the consequences of the decision are important. In situations of this type decomposition provides a tool whereby the decision maker can organize his thinking and, through sensitivity testing, determine which of the attributes really are crucial to his final decision.

Finally, decomposition is better suited than bootstrapping for application in organizations where the decision process involves many participants. As Edwards (1971) has argued, decomposition facilitates division of labor, with specialists making assessments within their own area of expertise, and decision makers with overall responsibility assigning weighting factors to the component attributes. The methodology can also serve to improve interpersonal communications involved in the decision process. Differences of opinion will be directly related to different interpretations of what the parameters of the evaluation model should be, and resolution of these differences can be accomplished by discussing those specific aspects of the evaluation model which do in fact produce the disagreement. Edwards (1971) discusses a case where decision maker's thought that they disagreed about the optimal course of action. But when decomposition models were constructed it was found that though they disagreed about the relative importance of certain attributes, the resulting models favored the same course of action for both decision makers.

Riskless Choice and the Additivity Assumption

As Edwards and Tversky (1967) noted, additive models have dominated discussions of riskless choice at least in large part because of their mathematical simplicity. Their use has been further reinforced by the degree to which additive statistical models have been able to approximate the human judgment process. But, when one's goal is to assist the decision maker, and to improve the quality of his decisions, then it becomes necessary to evaluate the appropriateness of the additive form.

The theory of conjoint measurement (Adams and Fagot, 1960; Luce and Tukey, 1964; Krantz and Tversky, 1970; Krantz, Luce, Suppes, and Tversky, 1971) provides a formal axiomatic basis for additivity. The theory requires that preferences be weakly ordered; that is, that preferences satisfy the following two properties:

1. Connectedness: for any two outcomes X and Y, either $X \precsim Y$, $Y \precsim X$, or both.
2. Transitivity: for any three outcomes X, Y, and Z, if $X \precsim Y$ and $Y \precsim Z$, then $X \precsim Z$.

For a finite set of outcomes, the weak ordering property alone is sufficient to guarantee the existence of some (not necessarily additive) value function V such that, for any X and Y: $X \precsim Y$ if and only $V(X) \leq V(Y)$. Further, if V is a value function, then any monotone transform of V is also a value function. Historically, such functions have been referred to as "riskless utility functions" (Luce and Suppes, 1965); throughout this paper, however, we have followed Raiffa's (1969) convention of using the term "value" in a riskless choice context, and "utility" in a risky choice setting.

As noted above, the weak ordering property guarantees the existence of some value function V, but places no restrictions on the functional form of V. Additional assumptions are required to guarantee that V will be an additive function of outcome attributes. For the two attribute outcome case, the major empirically testable requirement is that the double cancellation property be satisfied. Let X_1, X_2, and X_3 be any outcomes on the first attribute X, and let Y_1, Y_2, and Y_3 be any outcomes on the second attribute Y. Double cancellation requires that for all $X_1, X_2, X_3, Y_1, Y_2, Y_3$, if $(X_1,Y_3) \precsim (X_3,Y_2)$ and $(X_3,Y_1) \precsim (X_2,Y_3)$, then $(X_1,Y_1) \precsim (X_2,Y_2)$. Additivity requires further technical assumptions about the denseness of outcomes along component dimensions. But, for practical purposes, satisfaction of cancellation is usually taken as sufficient evidence for the existence of an additive value function V comprised of component functions V_1 and V_2 such that for any two outcomes $X=(X_1,X_2)$ and $Y=(Y_1,Y_2)$, $X \precsim Y$ if and only if $V_1(X_1) + V_2(X_2) \leqq V_1(Y_1) + V_2(Y_2)$. Further, when the other axioms are satisfied, V_1 and V_2 are defined on an interval scale.

When outcomes are characterized by three or more attributes, the cancellation axiom is replaced by the more intuitive monotonicity condition discussed earlier. As before, let the n-attributed outcome X^i be denoted by the vector $(X_1^i,X_2^i,\ldots,X_n^i)$. Let Y^i be any subset of these attributes, and let Z^i be the vector of attributes not contained in Y^i. Then monotonicity requires that preferences over the Y attributes, holding the Z attributes at some arbitrary levels, be unaffected by the particular levels at which the Z attributes are fixed. That is, for all possible partitions of X into Y and Z subsets, and for any Y^i and Y^j, if for any Z', $(Y^i,Z') \precsim (Y^j,Z')$, then for all Z", $(Y^i,Z'') \precsim (Y^j,Z'')$. When weak ordering, monotonicity, and the technical assumptions are satisfied then there will exist interval scale constituent functions $V_1,V_2,\ldots,V_n$ such that for any two outcomes X^i and X^j, $X^i \precsim X^j$ if and only if $\Sigma_k V_k(X_k^i) \leqq \Sigma_k V_k(X_k^j)$.

To summarize, theories of conjoint measurement provide axiom systems guaranteeing the existence of an interval scale additive value function. Descriptively, conjoint measurement may be used in place of the analysis of variance for testing the additivity assumption and for scaling value functions. As a normative tool, conjoint measurement is primarily useful for the insights it provides into the additivity assumption. When cancellation and monotonicity are viewed as appropriate by the decision maker, then an appropriate additive evaluation scheme can be constructed. But when these assumptions are not valid, then other evaluation procedures are called for. As Edwards (1971) has noted, however, it is difficult to imagine circumstances for which monotonicity does not hold. And when it does not, it will often be possible to restructure the attributes so that it does.

But despite the extreme generality of additivity in the conjoint measurement sense, additive statistical models are not quite so robust as has been commonly asserted in the judgment literature. Conjoint measurement views both the dependent and independent variables as having only ordinal properties and obtains an additive representation by rescaling both. Analysis of variance, on the other hand, permits non-linear transformations of the independent variables only, and linear regression permits only linear transformations. Thus, the statistical models which have been typically applied to experimental data are considerably more restrictive than the conjoint measurement formulation. Nevertheless, numerical examples presented by Yntema and Torgerson (1961) and Pollack (1964) have led to the erroneous conclusion that even with an interval scale dependent variable, satisfaction of monotonicity is sufficient to guarantee a good fit for additive statistical models. This argument has been widely offered as an explanation for why regression studies have found no evidence of non-additive judgment processes.

The following numerical examples reveal, however, that additive statistical models are not capable of approximating all simple functions which satisfy the monotonicity condition (Fischer, in prep.). Only two functions were considered. The first, F_1, consists of additive terms plus all possible two way multiplicative interactions.

$$F_1(X_1, X_2, \ldots, X_n) = \Sigma X_i + \sum_{i \neq j} \sum X_i X_j$$

The second function, F_2, combines the additive terms with a product over all terms.

$$F_2(X_1, X_2, \ldots, X_n) = \Sigma X_i + X_1 X_2 \ldots X_n$$

For the present analysis, examples with two, four, and six attributes were considered. Each attribute could assume integer values from one to ten. For each example, samples of size 1000 were used, with attribute values being randomly generated from a uniform distribution over the integers 1,2,...,10. To assess the ability of simple additive main effects models to approximate F_1 and F_2 a correlational analysis was conducted. The results of this analysis are summarized below. Cell entries give the correlation of an additive model with the non-additive function in question.

Number of Attributes	F_1	F_2
2	.955	.955
4	.978	.810
6	.987	.707

The additive approximation gives a good account to F_1 which involves only two way interactions, and the approximation improves with the number of attributes. The additive approximation to F_2, on the other hand, becomes poor with an increasing number of attributes.

From a normative standpoint, these results present no problem. If cancellation and monotonicity are viewed as appropriate then it is possible to construct some additive value function which will reflect the ordinal properties of the decision maker's preferences. Since riskless choice depends only upon these ordinal properties a function so constructed is appropriate for normative application.

From a descriptive standpoint, however, the numerical results presented above require a reinterpretation of the results of a number of judgment studies. In particular, studies using an interval scale dependent variable and a linear statistical analysis have frequently failed to obtain evidence of substantial departures from additivity. This failure is often attributed to the robustness of additive models, with references to the Yntema and Torgerson example (for example, Slovic and Lichtenstein, 1971). The numerical results presented above, however, demonstrate that if one takes seriously the interval scale properties of his dependent variable, then an additive approximation can give a good fit to only a limited subset of the models satisfying the monotonicity condition. As a consequence, the success of linear statistical models in explaining data may be considerably more informative about the underlying psychological processes than has generally been realized. Anderson (1970) describes a number of procedures which are useful for extracting such information.

MULTI-DIMENSIONAL UTILITY ASSESSMENT UNDER RISK

The Decision Theory Approach

Choice is said to be risky when the decision maker is unsure of the consequences which will result from each possible course of action, but is able to express this uncertainty in the form of probability distributions over outcomes. The term risky has frequently been restricted to decision making contexts for which some "objective probability" measure is defined. A growing number of decision theorists, however, are now adopting the Bayesian approach which views probability as a measure of the decision maker's knowledge or beliefs about states of the world (Ramsey, 1931; Savage, 1954; Kyburg and Smokler, 1964; Raiffa, 1968). Throughout this discussion we will adopt the position that some (possibly subjective) probability measure is available, and thus, that the decisions to be made can be appropriately viewed as risky.

Decision theory usually treats the decision maker's (DM's) uncertainty about the consequences of actions as arising from uncertainty about states of nature or the world. Let $(A^1,A^2,\ldots,A^n)$ be the set of actions available to DM, and let $(S^1,S^2,\ldots,S^r)$ be a mutually exclusive and exhaustive set of states of the world. For example, the action alternatives might be "Carry an umbrella to work" and "Don't carry an umbrella to work", and the states of the world "Will rain" and "Won't rain". Each act-state pair defines a consequence or outcome. Notationally, let X^{ij} denote the outcome arising from the $\underline{i}$-th act and the $\underline{j}$-th state of the world. In the umbrella example, the outcomes are "Carry umbrella; Rain", "Carry umbrella; No rain", "Don't carry umbrella; Rain", and "Don't carry umbrella; No rain".

Each course of action can result in one and only one outcome. But at the time when he must make his decision, DM is unsure of what that outcome will be. By assumption, however, he can specify a set of probabilities $p_1, p_2, \ldots, p_r$, where $\underline{p_j}$ is the probability of state S^j and where $\sum_j p_j = 1$. So each action A^i may be viewed as giving rise to a probability distribution or lottery over possible consequences. Notationally, let L^i be lottery of consequences associated with A^i, where

$$L^i = (p_1, X^{i1}; p_2, X^{i2}; \ldots ; p_r, X^{ir}).$$

That is, if action A^i is chosen, outcome X^{i1} will occur with probability $\underline{p_1}$, outcome X^{i2} with probability $\underline{p_2}$, and so on. The problem confronting DM is to select that course of action whose associated lottery of consequences is most desirable.

A number of procedures for choosing between probability distributions of outcomes have been proposed (Luce and Raiffa, 1957), but the expected utility principle has dominated normative discussions of the risky choice problem. According to this principle there exists some interval scale utility function U such that DM should choose that course of action for which the associated expected utility is greatest. Given a utility function defined on outcomes, the expected utility for action A^i is given by

$$EU(A^i) = p_1 U(X^{i1}) + p_2 U(X^{i2}) + \ldots + p_r U(X^{ir}).$$

The expected utility principle is not new; Bernoulli discussed it during the 1700's. But its prominence as a normative principle was greatly enhanced when von Neumann and Morgenstern (1944) showed that it could be deduced from a set of basic principles of rational choice, such as the transitivity and weak ordering properties discussed earlier. Other axiom systems for expected utility theory have since been developed (Herstein and Milnor, 1953;Savage, 1954; Luce and Raiffa, 1957). Elementary presentations of expected utility theory may be found in Raiffa (1968), Coombs, Dawes, and Tversky (1970), or Lee (1971), and more formal treatments in Luce and Raiffa (1957) and Fishburn (1970).

The Distinction Between Value and Utility.

As noted earlier, decision theorists frequently distinguish between risky and riskless utility functions. We, however, have adopted the convention of restricting use of the term utility to risky choice contexts, and have used the term value in riskless contexts. Technically, value and utility functions may be defined as follows (Fishburn, 1968):

1) A function V defined on a set of outcomes $(X^1,X^2,...,X^n)$ is said to be a value function whenever for any X^i and X^j in the outcome set

$$X^i \precsim X^j \text{ if and only if } V(X^i) \leqq V(X^j).$$

Further, if V is a value function, then any monotonic transform of V is also a value function.

2) A function U defined on a set of outcomes $(X^1,X^2,...,X^n)$ is said to be a utility function whenever for any X^i and X^j in the outcome set

a) $X^i \precsim X^j$ if and only if $U(X^i) \leqq U(X^j)$,

b) and $U(p,X^i;1-p;X^j) = pU(X^i) + (1-p)U(X^j)$.

And if U is a utility function, then any positive linear transform of U is also a utility function.

Thus, value and utility functions differ in their uniqueness properties. If DM maximizes the expectation of U under risk, then it is also appropriate that he maximize U in the absence of risk. But if DM maximizes V in the absence of risk, it does not follow that he should maximize the expectation of V under risk. For example, most people prefer more money to less, and so (all other value factors being held constant) may be viewed as maximizers of monetary return in the absence of risk. But in general people do not act as maximizers of expected monetary return under risk; they buy insurance, for example, even though the expected monetary return of such a purchase is negative. If, however, the DM is an expected utility maximizer, then there will exist some transformation on dollars, U($), such that DM will maximize the expectation of U($) under risk. Thus, utility may in this case be viewed as a transformation on dollars reflecting DM's attitude toward risk. Raiffa (1968) provides an elementary discussion of the utility theory treatment of aversion, or attraction, to risk.

A similar argument can be made in the context of multi-attribute preference. Though an additive function V may be an appropriate measure of value in a riskless context, it does not follow that DM should maximize the expectation of V in the presence of risk (Raiffa, 1969).

<u>Utility for Multi-attribute Outcomes</u>.

In principle utility theory is applicable to any class of outcomes. But in practice utility theory applications have generally been restricted to the single attribute case. Recently, however, utility theorists have devoted considerable attention to the properties of various utility functions defined on multi-attribute outcomes. As in the riskless case, additive functions have received extensive consideration. Assuming the existence of some function U satisfying the utility theory axioms, Fishburn (1964) has specified an additional assumption which guarantees that U will be additive. Central to Fishburn's development is a relationship between finite gambles which we will term marginal equivalence. Two gambles (or lotteries) L^1 and L^2 are marginally equivalent if and only if they give rise to identical marginal probability distributions over outcome attributes. This relation can easily be illustrated for the case of two attribute outcomes. Let L^1 be (1/2,(X",Y");1/2,(X',Y')) and L^2 be (1/2,(X",Y');1/2,(X',Y")). For both lotteries the probability of obtaining attribute X" is 1/2 and the probability of obtaining attribute X' is 1/2; similarly, the probabilities of obtaining attributes Y" and Y' are also 1/2 for both lotteries. Thus, the two lotteries are marginally equivalent. Using this definition we can now state the Fishburn marginality assumption - for all lotteries L^i and L^j, if L^i and L^j are marginally equivalent, then DM is indifferent between L^i and L^j. Given the assumed existence of some utility function U defined on a set of multi-attribute outcomes, Fishburn has shown that U is additive if and only if the Fishburn marginality assumption is satisfied.

Thus it is possible to assess the desirability of the additive formulation by examining the implications of the marginality assumption. For example, consider the following pair of marginally equivalent lotteries:

$$L^1 = \begin{cases} \text{with probability } 1/2\text{, receive \$5000 and a 1972 Volvo} \\ \text{with probability } 1/2\text{, receive \$10 and a rusty hubcap} \end{cases}$$

$$L^2 = \begin{cases} \text{with probability } 1/2\text{, receive \$5000 and a rusty hubcap} \\ \text{with probability } 1/2\text{, receive \$10 and a 1972 Volvo.} \end{cases}$$

An additive utility function defined on dollars and automobile components exists if and only if the decision maker is indifferent between L^1 and L^2. A casual survey indicates that most people are not indifferent. They prefer L^2 which provides a sure thing of obtaining either \$5000 or a Volvo. In general, it seems that there will be few circumstances for which the marginality assumption will be satisfied. And when it is not, non-additive utility functions will be required.

Keeney (1969,1971) has discussed a special class of non-additive utility functions which arise when the <u>mutual utility independence</u> condition (defined below) is satisfied. Let X_j be the <u>j</u>-th attribute of the generic outcome $(X_1, X_2, \ldots, X_n)$, and let $\bar{X}_j$ be the vector of remaining attributes for this generic outcome. Then Keeney's mutual utility independence condition requires that for all $j = 1,2,\ldots,n$,

$$U(X_j,\bar{X}_j') = c_1(\bar{X}_j') + c_2(\bar{X}_j')\, U(X_j,\bar{X}_j^o),\quad c_2>0.$$

That is, utility for X_j conditional upon $\bar{X}_j'$ is a positive linear transformation of utility for X_j conditional upon any other $\bar{X}_j^o$. Keeney showed that whenever the utility theory axioms are satisfied and mutual utility independence is also satisfied, then the utility function U will consist only of additive and multiplicative components. For example, in the three dimensional case there will exist conditional utility functions U_1, U_2, and U_3

such that

$$U(X_1,X_2,X_3) = U_1(X_1) + U_2(X_2) + U_3(X_3) + k_1U_1(X_1)U_2(X_2)$$
$$+ k_2U_1(X_1)U_3(X_3) + k_3U_2(X_2)U_3(X_3) + k_4U_1(X_1)U_2(X_2)U_3(X_3),$$

where k_1, k_2, k_3, and k_4 are constants to be empirically determined. Keeney terms functions of this type <u>quasi-additive</u>. It can easily be shown that the existence of a quasi-additive risky utility function implies the existence of an additive value function, since mutual utility independence implies both cancellation and monotonicity.

Raiffa (1969) has discussed an even less restrictive class of non-additive utility functions. Given the existence of an additive value function V and given that the utility theory axioms are satisfied, then by the definition of value and utility functions there exists some monotone transform Φ such that

$$U(X_1,X_2,\ldots,X_n) = \Phi(V(X_1,X_2,\ldots,X_n))$$
$$= \Phi(V_1(X_1) + V_2(X_2) + \ldots + V_n(X_n)).$$

Methods for assessing the functional form of Φ will be discussed in a subsequent section of this paper.

<u>Utility Theory as a Descriptive Model of Human Decision Making</u>.

Though derived from normative principles, utility theory has also been applied descriptively. Reviews of the relevant literature may be found in Edwards (1961), Luce and Suppes (1965), and Becker and McClintock (1967). In general, the expected utility hypothesis has been tested only for decisions involving choices between simple gambles (Mosteller and Nogee, 1951; Davidson, Suppes, and Siegel, 1957; Coombs, Bezembinder, and Goode, 1965). To illustrate these gambling studies we will consider the results of an experiment conducted by Tversky (1967). This study is outstanding in two regards. First, it provides the most direct test of the expected utility hypothesis. Second, it is the only study to incorporate multi-attribute outcomes. Subjects were presented with gambles of the form $(p;x,y)$, to be interpreted as "With probability p you will receive x packs of cigarettes and y bags of candy; with probability $1-p$ you will receive nothing." Subjects were then asked to state minimum selling prices for these gambles. Assuming that the subjects were utility maximizers, and arbitrarily letting the utility of receiving nothing be zero, we have

$$U(MSP) = S(p)\ U(x,y),$$

where MSP is the minimum selling price for the gamble in question, and $S(p)$ is the subjective probability of winning given objective probability p of winning. Taking logs of both sides of this equation

$$\log U(MSP) = \log S(p) + \log U(x,y).$$

As a consequence, the expected utility hypothesis can be satisfied if and only if probabilities and payoffs combine additively in the conjoint measurement sense. Tversky's data provided strong support for the expected utility hypothesis. In addition, he obtained utility functions for single dimensional outcomes involving either cigarettes or candy and, assuming that utilities for two dimensional outcomes were additive, used these functions to predict the selling prices for the two attribute bundles of cigarettes and candy. The accuracy of these predictions indicated that in this case utilities did combine additively over attributes in the presence of risk. In general, Tversky's study provides strong support for an expected utility interpretation of decision making under risk. A more recent study conducted by Goodman, Saltzman, Edwards, and Krantz (1971) further attests to the predictive power of expectation models. They found that a simple expected monetary value maximization model accounted for almost all of the systematic variance in gambling behavior involving fairly sizable outcomes.

Lichtenstein and Slovic (1971), however, have questioned whether expected utility models do adequately represent even simple decision making processes. They found that different response modes led to different preference orderings on gambles, a result inconsistent with the utility theory axioms. They argued that descriptive models of human choice must take into account cognitive factors which are ignored by utility models. Nevertheless, the results cited above leave little doubt that utility theory provides a good first approximation to decision making under risk, at least for simple gambling situations. This conclusion will be seen to be crucial for those who wish to apply utility theory as a no mative procedure for making real world decisions.

Multi-attribute Utility Assessment: Aiding the Decision-maker.

The most direct normative applications of the expected utility theory principles have been in the field of mathematical statistics (Savage, 1954; Chernoff and Moses, 1959; Raiffa and Schlaiffer, 1961; DeGroot, 1970). In addition, the discipline of decision analysis has arisen. Decision analysis is a set of procedures for combining the expected utility principle with a Bayesian interpretation of probability for the purpose of making real world decisions (Edwards, Lindman, and Phillips, 1965; Raiffa, 1968; Howard, 1968a, Howard, 1968b). In the past, decision analysts were primarily concerned with the problem of probability assessment (for example, Edwards, Phillips, Hays, and Goodman, 1968). Only recently has it become apparent that utility assessment procedures are insufficiently developed for application in many real world contexts. In this section we will consider first the general logic of utility assessment, then practical procedures for coping with the additional complexities introduced by multi-attribute outcomes.

The Logic of Utility Assessment. The decision analysis approach rests on the assumption that, when faced with complex problems and left to their own devices, decision makers do not act in an expected utility maximizing fashion. The formal methods of decision analysis are tools for producing more nearly optimal decisions. Yet, because decision analysts justify their approach by arguing the normative merits of the utility theory axioms a mild paradox arises. For a proper utility function can be assessed only if the judgments upon which it is based are made in an expected utility maximizing fashion. Since utility functions are typically inferred from judgments about simple gambles, the decision analyst must in effect assume the decision maker is an expected utility maximizer for simple decisions, but is sub-optimal when faced with complex decisions. This does not seem an unrealistic assumption.

Although a number of procedures for assessing risky utilities have been developed, they utilize a common logic. To illustrate this logic we will consider Raiffa's (1968) indifference probability procedure. Consider a set of outcomes $(X^1, X^2, \ldots, X^n)$. Let X^* and X_* be the most and least preferred elements of this set, and arbitrarily assign these outcomes utilities of 1.0 and 0.0 respectively. Then the utility of any other outcome X^i may be obtained as follows. The utility theory axioms assert that for any outcome X^i there exists some probability $\underline{p}^i$ such that DM is indifferent between X^i and the lottery $(p^i, X^*; 1-p^i, X_*)$. So from $X^i \sim (p^i, X^*; 1-p^i, X_*)$ we have $U(X^i)=p^i U(X^*)+(1-p^i)U(X_*)$. Or $U(X^i)=p^i$.

Since the procedure is neutral with respect to the composition of outcomes it can in principle be used for either single or multi-attributed outcomes. In practice, however, the method is difficult to apply in the multi-attribute case. To ask decision makers to simultaneously aggregate value and risk over ten or more attributes seems unreasonable, unless one is willing to tolerate the possibility of a substantial degree of unreliability. And even when the DM is able to make reliable judgments, the time required for evaluating a large number of alternatives may in many cases be prohibitive.

To offset these problems, decomposition procedures for risky utilities have been proposed. Central to these procedures is the assumption that DM can make meaningful probabilistic utility assessments within a given dimension. In addition, some of the procedures require DM to make "a few" such assessments across dimensions.

Before proceeding further, we will briefly consider some results obtained by Ginsberg (1969) which strongly indicate that probabilistic utility assessments can be made in a reliable fashion. Ginsberg was concerned with the problem of scaling the utility of severe outcomes (such as loss of sight or limbs) which arise in the course of medical practice. Three trained physicians assigned utilities to eight such outcomes using the indifference probability method described above. In addition, they directly estimated the dollar amount which they themselves would pay in order to avert each of these eight dire outcomes. Finally, each physician assessed a utility function for money. The set of scaling methods used permitted Ginsberg to compute the direct dollar judgments implied by the indifference probability judgments. The correlations between actual dollar bids and predicted dollar bids were remarkably high; .997, .983, and .998 for the three doctors respectively. This high degree of convergence indicates that the indifference probability judgment task did "make sense" to the doctors, and that they could respond to it in a meaningful fashion.

Decomposed Utility Assessment. Three general methods for obtaining a decomposed risky utility function have been proposed. These methods are based on the additive, quasi-additive, and Φ (V) utility models, respectively. Recall that all three models assume the existence of some function U satisfying the standard utility theory exioms. The models differ in terms of how component attributes contribute to overall utility. The additive model asserts that the overall utility of a multi-attribute outcome is an additive function of the utilities of its component attributes. For example,

$$U(X_1,X_2) = U_1(X_1) + U_2(X_2).$$

In the quasi-additive model, overall utility is a function of both additive and multiplicative cross-product terms. For example,

$$U(X_1,X_2) = U_1(X_1) + U_2(X_2) + kU_1(X_1)U_2(X_2).$$

Finally, the $\Phi(V)$ model simply asserts that overall utility is a monotonic function of overall riskless value. For example, if V is an additive value function such that $V(X_1,X_2)=V_1(X_1)+V_2(X_2)$, then this model assumes the existence of some monotone transform Φ such that $U(X_1,X_2)= \Phi(V_1[X_1]+V_2[X_2])$. In this section we will consider procedures for obtaining decomposed utility functions of the three forms described above. In the final section we will discuss experimental attempts to validate these decomposition procedures.

As noted earlier, the additive form is appropriate only when the expected utility, monotonicity, and marginality assumptions are satisfied. When this is the case, the following procedure devised by Raiffa (1969) may be used to obtain an additive utility function. Like the rating scale procedure for riskless choice, Raiffa's method involves four major steps.

1. Within each attribute, DM must apecify the most and least desirable outcomes which may feasibly occur. Notationally, let X_i^* and X_{i*} denote the most and least desirable outcomes with respect to the i-th attribute. Arbitrarily, we assign utilities of 1.0 and 0.0 to these two outcomes respectively.

2. DM must next assign utilities to intermediately valued outcomes on each attribute. Again, let $(X_1^o,X_2^o,\ldots,X_n^o)$ be the vector of the "standard outcome". Then for each attribute, DM can assess a utility function over the possible outcomes on this attribute assuming all other attribues to be held constant at their standard values. These utility functions are obtained using the indifference probability method.

3. Cross dimensional scaling is accomplished as follows. Let $X^*=(X_1^*,X_2^*,\ldots,X_n^*)$ and $X_*=(X_{1*},X_{2*},\ldots,X_{n*})$ be the best and worst multi-attributed outcomes, and arbitrarily assign them overall utilities of 1.0 and 0.0 respectively. Let $(X_i^*,\overline{X}_{i*})$ denote the consequence which has the best feasible outcome with respect to the i-th attribute and the worst feasible outcome with respect to all other attribues. For each attribute DM must specify a probability π^i such that he is indifferent between $(X_i^*,\overline{X}_{i*})$ and the lottery $(\pi^i,X^*;1-\pi^i,X_*)$. It can be shown that π^i is a measure of the utility range associated with the i-th attribute. Under the assumption that U is additve, $\sum_i \pi^i = 1.0$, so the untransformed π^i may be used as scaling factors.

4. The utility of any multi-attributed outcome is thus given by $U(X_1^k,X_2^k,\ldots,X_n^k) = \sum_i \pi^i U_i(X_i^k)$. When a decision is to be made, that action with the greatest associated expected utility is selected.

Given the marked similarity between this method and the additive rating scale procedure for riskless choice, it might seem that the two methods should be interchangeable. Raiffa (1969) has shown, however, that even when a risky additive utility function exists, it does not necessarily follow that a bona fide riskless additive value function will be appropriate in a risky context. On the other hand, if an additive utility function is appropriate in a risky context, then it is also appropriate in a riskless context. Intuitively, the probabilistic procedure described above reflects attitude toward risk within each attribute whereas the rating scale method does not.

A quasi-additive utility function involves additive and multiplicative cross product terms, and arises when the expected utility, monotonicity, and mutual utility independence conditions are satisfied, but the marginality assumption is not. Given that a quasi-additive form is appropriate, the following decomposition procedure may be employed (Keeney, 1969; Raiffa, 1969).

1. Utility functions within dimensions may be obtained as in the additive case.

2. In order to establish a common scale of measure across attributes DM must intuitively assess, for each attribute, the utilities of $(X_1^o, X_2^o, \ldots, X_{i-1}^o, X_i^*, X_{i+1}^o, \ldots, X_n^o)$ and $(X_1^0, X_2^o, \ldots, X_{i-1}^o, X_{i*}, X_{i+1}^o, \ldots, X_n^o)$. These judgments determine the utility range associated with each of the dimensions, given that all other dimensions are held at their standard levels. (In contrast to the additive model, utility ranges for the quasi-additive model depend upon the state of other attributes.)

3. Weighting factors for the multiplicative terms of the quasi-additive model are obtained by having the DM intuitively assess the utility of all the "corner points" in the outcome space. For example, with three attributes there are eight corner points: (X_{1*}, X_{2*}, X_{3*}), (X_1^*, X_{2*}, X_{3*}), (X_{1*}, X_2^*, X_{3*}), (X_{1*}, X_{2*}, X_3^*), (X_1^*, X_2^*, X_{3*}), (X_1^*, X_{2*}, X_3^*), (X_{1*}, X_2^*, X_3^*), and (X_1^*, X_2^*, X_3^*). In general, there will be 2^n corner points, where $\underline{n}$ is the number of attributes. Keeney (1969) provides formulas for using these corner point assessments for weighting the cross product terms.

The $\Phi(V)$ procedure is the most general of the decomposition methods. It requires only that the expected utility and monotonicity assumptions be appropriate. And despite its generality, it may well be the easiest utility assessment procedure to implement. In the first stage a riskless value function must be developed. Either the additive rating scale or trade-off procedures may be used. When the trade-off method is employed it may be quite simple to obtain the desired risk transformation Φ. Suppose, for example, that all outcomes have been traded off into a single continuous dimension such as dollars or lives saved. Then Φ may be obtained by assessing a unidimensional risky utility function over this continuous attribute.

It is also possible to obtain Φ when a rating scale decomposition is used. Here the decision maker is required to directly assess the utility of a few well chosen multi-attribute outcomes (Raiffa, 1969). The values of these outcomes (as indicated by the additive rating scale model) are then plotted on one axis, and the utilities of these outcomes on the other axis. Utilities for outcomes having other values can be obtained by interpolating a smooth curve through the selected points for which utilities have been assessed.

In working with subjects the author has observed that the Φ(Value) approach is quite easy to implement because it requires few probabilistic judgments of the subject. Moreover, because this method requires the least restrictive assumptions, it can be appropriately utilized when the additional assumptions required of the additive or quasi-additive methods are also satisfied.

Experimental Validation of Decomposed Utility Procedures

To date only three validation studies have been conducted, and all three have used convergence between utility measures as the validating criterion. Von Winterfeldt (1971) had subjects evaluate the attractiveness of apartments described by fourteen attributes. In the first stage of the experiment subjects assigned overall intuitive utilities to a set of hypothetical apartments using Raiffa's indifference probability procedure. Next, additive decomposed risky utility functions were assessed by each subject using the method discussed in the previous section. Finally, intuitive overall utilities were reassessed. The mean correlation between the decomposed utilities and the second set of intuitive utility judgments was .84.

In a similar study Fischer (in prep.) had subjects assign utilities to hypothetical compact cars described by either three or nine attributes. For three dimensions the convergence between intuitive and additive decomposed utilities was quite high (median $R=.93$); but for nine dimensions convergence dropped off slightly (median $R=.85$).

Finally, Fischer (in prep.) has contrasted the predictive power of additive utility decompositions with that of $\Phi(V)$ decompositions. The $\Phi(V)$ method could be expected to be superior under either of two circumstances. First, if intuitive utility assessments are systematically non-additive then the $\Phi(V)$ method, which can capture this non-additivity, should outperform the additive utility decomposition, which cannot. Second, even if intuitive utility assessments are additive, the $\Phi(V)$ method is still appropriate and might produce less judgmental error, thus yielding superior predictions of intuitive judgments.

In Fischer's second study, subjects evaluated jobs described by three attributes--city, salary, and type of work. First, overall intuitive utilities were assessed, then additive and $\Phi(V)$ decompositions constructed. Each subject assigned utilities to all 27 combinations of the three attributes, thus permitting a direct test of the hypothesis that intuitive multi-attribute preferences under risk are additive. Next, each subject made the judgments required for constructing additive and $\Phi(V)$ decomposition models. The results of this study strongly indicated that an additive formulation was adequate. An across subjects analysis of variance was performed on the intuitive utility judgments. Additive main effects accounted for 98.8% of the final effects sums of squares. In addition, the additive and $\Phi(V)$ decompositions provided essentially equal prediction of the intuitive judgments, with median correlations of .925 and .935, respectively. Finally, a reliability analysis indicated that the degree of prediction afforded by the decomposition models approached the limits set by the error variance in the intuitive judgments.

Nevertheless, risky multi-attribute utility assessment deserves considerably greater attention. The use of intuitive judgments as a validating device here is clearly subject to criticism. In the riskless case, the additive formulation is consistent with basic normative assumptions. But, in a risky context the additive form requires questionable assumptions. And it may be the case that intuitive judgments are additive simply because the decision maker is unable to subjectively process information in a more complex fashion. If this were the case, then non-additive forms might be preferred on normative grounds even if they afforded poorer prediction of intuitive judgments. These questions can be resolved only through additional research.

References

Adams, E. W., and Fagot, R. A model of riskless choice. Behavioral Science, 1959, 4, 1-10. Reprinted in W. Edwards and A. Tversky (Eds.), Decision Making, Penguin, 1967.

Anderson, N. H. Application of an additive model to impression formation. Science, 1962, 138, 817-818.

Anderson, N. H. Averaging model analysis of set size effects in impression formation. Journal of Experimental Psychology, 1967, 75, 158-165.

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

Anderson, N. H., and Jacobson, A. Effects of stimulus inconsistency and discounting instructions in personality impression formation. Journal of Personality and Social Psychology, 1965, 2, 531-539.

Becker, G. M., and McClintock, C. M. Value: behavioral decision theory. Annual Review of Psychology, 1967, 18, 239-286.

Bowman, E. H. Consistency and optimality in managerial decision making, Management Science, 1963, 9, 310-321.

Chernoff, H., and Moses, L. Elementary Decision Theory, New York: Wiley, 1959.

Coombs, C. H., Bezembinder, T. G. G., and Goode, G. M. Testing expectation theories of decision making without measuring utility or subjective probability: a supplementary report. The University of Michigan Mathematical Psychology Program, MMPP 65-5, October, 1965.

Coombs, C. H., Dawes, R. M., and Tversky, A. Mathematical Psychology: An Elementary Introduction, Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Crecine, J. P, and Fischer, G. W. On resource allocation processes in the U. S. Department of Defense. The University of Michigan, Institute of Public Policy Studies Discussion Paper No. 31, 1971.

Cyert, R. M., and March, J. G. A Behavioral Theory of the Firm, Englewood Cliffs, N. J.: Prentice-Hall, 1963.

Davidson, D., Suppes, P., and Siegel, S. Decision-making: An Experimental Approach, Stanford University Press, 1957.

Davis, O. A., Dempster, M. A. H., and Wildavsky, A. On the process of budgeting II: an empirical study of Congressional appropriations. In R. F. Byrne, et al. (Eds.), Studies in Budgeting, Amsterdam: North Holland Publishing, 1971.

Dawes, R. M. Social selection based on multi-dimensional criteria. Journal of Abnormal and Social Psychology, 1964, 68, 104-109.

Dawes, R. M. Graduate admissions: a case study. Oregon Research Institute Technical Report, 10(1), 1970.

De Groot, M. H., Optimal Statistical Decisions, New York: McGraw Hill, 1970.

Eckenrode, R. T. Weighting multiple criteria. Management Science, 1965, 12, #3.

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

Edwards, W. Social utilities. Paper presented at the Symposium on Decision and Risk Analysis, American Society for Engineering Education and American Institute of Industrial Engineers, Annapolis, Maryland, June, 1971, Printed by The University of Michigan Engineering Psychology Laboratory, 1971.

Edwards, W., Lindman, H., and Phillips, L. D. Emerging technologies for making decisions, in New Directions in Psychology II, New York: Holt, Rinehart, and Winston, 1965, 261-325.

Edwards, W., Phillips, L. D.,Hays, W., and Goodman, B. Probabilistic information processing systems: design and evaluation. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 248-265.

Edwards, W.,and Tversky, A. (Eds.) Decision Making, Penguin, 1967.

Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. Psychological Bulletin, 1970, 73,221-230.

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.

Fischer, G. W. Methods for assessing multi-attribute utilities: An empirical comparison. Proposed title for doctoral dissertation, The University of Michigan, in preparation.

Fishburn, P. C. Independence in utility theory with whole product sets. Operations Research, 1965, 13, 28-45.

Fishburn, P. C. Methods of estimating additive utilities. Management Science, 1967, 13, 435-453.

Fishburn, P. C. Utility theory. Management Science, 1968, 14, 335-378.

Fishburn, P. C. Utility Theory for Decision Making, New York: Wiley, 1970.

Fitts, P. M., and Posner, M. I. Human Performance, Belmont: Brooks/Cole, 1969.

Ginsberg, A. S. Decision analysis in clinical patient management: an application to the pleural effusion problem. Rand Memorandum 19753-RC, 1969.

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

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

Goodman, B., Saltzman, M., Edwards, W., and Krantz, D. H., When SEU is irrelevant. The University of Michigan, Department of Psychology, 1971.

Herstein, I. H., and Milnor, J. An axiomatic approach to measureable utility. Econometrica, 1953, 21, 291-297.

Himmelfarb, S., and Senn, D. J. Forming impressions of social class: two tests of an averaging model. Journal of Personality and Social Psychology, 1969, 12, 38-51.

Hoepfl, R. T., and Huber, G. P. A study of self explicated utility models. Behavioral Science, 1970, 15, 408-4[illegible].

Howard, R. A. The foundations of decision analysis. IEEE Transactions on Systems Science and Cybernetics, 1968a, SSC-4, 211-219.

Howard, R. A. Decision analysis: applied decision theory. In D. B. Hertz and J. Melese (Eds.), Proceedings of the Fourth International Conference on Operational Research, New York:Wiley, 1968b.

Huber, G. P., Daneshgar, R., and Ford, D. L. An empirical comparison of five utility models for predicting job preferences. Organizational Behavior and Human Performance, 1971, 6, 267-282.

Huber, G. P., Sahney, V. K., and Ford, D. L. A study of subjective evaluation models. Behavioral Science, 1969, 483-489.

Keeney, R. L. Multi-dimensional utility functions: theory, assessment, and application. Massachusetts Institute of Technology, Operations Research Center, Technical Report Number 43, October, 1969.

Keeney, R. L. Utility independence and preference for multi-attributed consequences. Operations Research, 1971, 19, #4.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. Foundations of Measurement: Additive and Polynomial Representations, I, New York: Academic Press, 1971.

Krantz, D. H., and Tversky, A. Conjoint-measurement analysis of composition rules in psychology. The University of Michigan Mathematical Psychology Program, MMPP 70-10, 1970.

Kyburg, H. E. Jr., and Smokler, H. E. (Eds.) Studies in Subjective Probabilities, New York: Wiley, 1964.

Lathrop, R. G., and Peters, B. E. Subjective cue weighting and decisions in a familiar task. Proceedings, 77-th Annual Convention, APA, 1969.

Lee, W. Decision Theory and Human Behavior, New York: Wiley, 1971.

Lichtenstein, S., and Slovic, P. Reversals of preferences between bids and choices in gambling decisions. Journal of Experimental Psychology, 1971, 89, 46-55.

Luce, R. D., and Raiffa, H. Games and Decisions: Introduction and Critical Survey, New York: Wiley, 1957.

Luce, R. D., and Suppes, P. Preference, utility, and subjective probability, In R. D. Luce et al. (Eds.), Handbook of Mathematical Psychology, II, New York: Wiley, 1965.

Luce, R. D., and Tukey, J. W. Simultaneous conjoint measurement: a new type of fundamental measurement. Journal of Mathematical Psychology, 1964, 1, 1-27.

MacKrimmon, K. R. Decision making among multiple-attribute alternatives: a survey and consolidated approach. The Rand Corporation, RM-4823-ARPA, 1968.

March, J. G., and Simon, H. Organizations, New York:Wiley, 1958.

Meehl, P. E. Clinical Vs. Statistical Prediction: A Theoretical Analysis and Review of the Evidence, The University of Minnesota Press, 1954.

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

Miller, J. R. III Assessing alternative transportation systems. Rand Memorandum RM-5865-DOT, April, 1969.

Miller, L. W., Kaplan, R. J., and Edwards, W. Judge: a value-judgment-based tactical command system. Organizational Behavior and Human Performance, 1967, 2, 329-374.

Miller, L. W., Kaplan, R. J., and Edwards, W. Judge: a laboratory evaluation. Organizational Behavior and Human Performance, 1969, 4, 97-111.

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

Norman, D. A. Memory and Attention, New York: Wiley, 1969.

O'Connor, M. F. The application of multi-attribute scaling techniques to the development of indices of water quality. Ph.D. Dissertation, The University of Michigan, Department of Psychology, 1972.

Pai, G. K., Gustafson, D. H., and Kiner, G. W. Comparison of three non-risk methods for determining a preference function. University of Wisconsin, January, 1971.

Pollack, I. Action selection and the Yntema-Torgerson worth function. In Information System Science and Engineering: Proceedings of the First Congress of the Information Systems Sciences, New York: McGraw-Hill, 1964.

Raiffa, H. Decision Analysis, Addison-Wesley, 1968.

Raiffa, H. Preferences for multi-attributed alternatives. The Rand Corporation, RM-5868-DOT/RC, April, 1969.

Raiffa, H., and Schlaiffer, R. Applied Statistical Decision Theory, Boston: Harvard Business School, 1961.

Ramsey, F. P. Truth and probability. In F. P. Ramsey, The Foundations of Mathematics and Other Logical Essays, New York: Harcourt Brace, 1931. Reprinted in H. E. Kyburg Jr., and H. E. Smokler (Eds.), Studies In Subjective Probabilities , New York: Wiley, 1964.

Savage, L. J. The Foundations of Statistics, New York: Wiley, 1954.

Sayeki, Y. Allocation of importance: an axiom system. University of Washington, Department of Psychology, July 1970.

Sayeki, Y., and Vesper, K. Allocation of importance and goal dependent utility. University of Washington, Department of Psychology, June 1971.

Shanteau, J. C., and Anderson, N. H. Test of a conflict model for preference judgment. Journal of Mathematical Psychology, 1969, 6,312-325.

Shepard, R. N. On subjectively optimum selection among multi-attribute alternatives. In M. W. Shelley and G. L. Bryan (Eds.), Human Judgment and Optimality, New York: Wiley, 1964. Reprinted in W. Edwards and A. Tversky (Eds.), Decision Making, Penguin, 1967.

Shepard, R. N., Hovland, C. I., and Jenkins, H. M. Learning and memorization of classifications, Psychological Monographs, 1961,75, (13, Whole, No. 517).

Sidowski, J. B. Judgments of city-occupation combinations. Psychonomic Science, 1967, 7, 279-280.

Simon, H. A. The Sciences of the Artificial, The M.I.T. Press, 1969.

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., and 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., Rorer, L. G., and Hoffman, P. J. Analyzing the use of diagnostic signs. Investigative Radiology, 1971, 6, 18-27.

Stigler, G. J. The Theory of Price, London: McMillan, 1966.

Toda, M. Indifference map method for estimating utility functions. Department of Psychology, Hokkaido University, Sapporo, Japan, 1971.

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

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

Tversky, A. Elimination by aspects: a probabilistic theory of choice. The University of Michigan Mathematical Psychology Program, MMPP 71-12, 1971.

Von Neumann, J., and Morgenstern, O. Theory of Games and Economic Behavior, Princeton University Press, 1944.

Von Winterfeldt, D. Multi-attribute utility theory: theoretical background and an experimental validation. Diploma Thesis, University of Hamburg (Germany), 1971.

Yntema, D. B., and Klem, L. Telling a computer how to evaluate multidimensional situations. IEEE Transactions on Human Factors in Electronics, 1965, HFE-6, 3-13.

Yntema, D. B., and Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transactions on Human Factors in Electronics, 1961, HFE-2, 20-26. Reprinted in W. Edwards and A. Tversky (Eds.), Decision Making, Penguin, 1967,