FOUR METHODS FOR ASSESSING MULTI-ATTRIBUTE UTILITIES: AN EXPERIMENTAL VALIDATION

Gregory W. Fischer

Michigan University

037230-6-T.

THE UNIVERSITY OF MICHIGAN

AD751564

ENGINEERING PSYCHOLOGY LABORATORY

Four Methods for Assessing Multi-Attribute Utilities: An Experimental Validation

Technical Report

GREGORY W. FISCHER

D D C

NOV 27 1972

B

Administered through:

September 1972

OFFICE OF RESEARCH ADMINISTRATION · ANN ARBOR

FOUR METHODS FOR ASSESSING MULTI-ATTRIBUTE UTILITIES:

AN EXPERIMENTAL VALIDATION

Technical Report

30 September 1972

Gregory W. Fischer

Engineering Psychology Laboratory

The University of Michigan

Ann Arbor, Michigan

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)	2a. REPORT SECURITY CLASSIFICATION
Department of Psychology University of Michigan Ann Arbor, Michigan	Unclassified
	2b. GROUP

3. REPORT TITLE

Four Methods for Assessing Multi-attribute Utilities: An Experimental Validation

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

Technical

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

Gregory W. Fischer

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
30 September 1972	90	33

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-67-A-0181-0034	037230-6-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

·In choosing between alternatives characterized by multiple value relevant attributes, decision makers must typically trade off one attribute against another. Previous research has shown that as the number of attributes describing alternatives becomes large, this subjective trade-off process becomes increasingly subject to error and that decision makers tend to ignore value relevant considerations. These shortcomings of the subjective evaluation process have been related to more general limitations on the human capacity to process information. Decomposed evaluation procedures seek to improve upon subjective evaluation by dividing the overall evaluation problem into a set of simpler subtasks, each of which is well within the judgmental capacities of the decision maker.

The first section of this paper discusses the theoretical basis for multi-attribute value assessment and concludes that while additive evelution models should be appropriate for most riskless decisions, non-additive models will frequently be required for decision making under uncertainty. The second section discusses the sensitivity of evaluation models to assessment errors. The third section describes four procedures for constructing a decomposed evaluation model. The fourth section treats the general problem of validating evaluation procedures. And the final two sections discuss two experiments which demonstrated that all four of the decomposition procedures described can provide an appropriate measure of subjective value.

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Multi-attribute Utility						
Riskless Choice						
Risky Choice						
Decision Making						
Value Function						
Utility Function						
Convergent Validation						

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF ILLUSTRATIONS

INTRODUCTION

Decision makers frequently choose between courses
of action whose probable consequences are characterized
by multiple value attributes. For example, corporate
strategies might be evaluated in terms of their impli-
cations for long run profits, short run profits, and
market share; automobiles in terms of their safety and
performance characteristics, cost, comfort, and luxury
options; and prospective graduate students in terms of
their test scores, recommendations, and undergraduate
academic records. Optimal decision making in such con-
texts requires that decision makers trade off one value
relevant factor against another in determining the overall
worth of each possible alternative. Economic theorists
(Edgeworth, 1881) have long assumed that people can sub-
jectively make such trade-offs, but until very recently
this assumption remained untested. During the past decade,
however, psychologists have devoted a substantial effort
to the study of the <u>wholistic</u> multi-attribute evaluation
process. (Throughout this paper an evaluation will be said
to be wholistic if it is generated by a subjective process
without resort to formal analytical procedures.) These

- 1 -

studies, reviewed by Slovic and Lichtenstein (1971), have generally indicated that people can make such trade-offs in a systematic and meaningful fashion. Nevertheless, this research has also revealed important shortcomings of wholistic evaluation. First, wholistic evaluations tend to be based on but a limited number of value attributes, frequently ignoring potentially significant value relevant considerations (Shepard, 1964; Slovic and Lichtenstein, 1971). This shortcoming seems to arise from a more general limitation of human information processing--namely, that people can deal with only five to ten "chunks" of conceptual information at any given time (Miller, 1956). In addition, wholistic evaluations are characterized by a substantial degree of random error, and the amount of error tends to increase as the decision maker attempts to consider an increasing number of value attributes (Slovic and Lichtenstein, 1971). Error of this type has been shown to be an important source of suboptimality in real world decision making (Bowman, 1963).

As the shortcomings of wholistic evaluation have become increasingly apparent, decomposed evaluation procedures have been proposed as means for improving upon the intuitive decision making process. Decomposition methods attempt to achieve greater optimality by dividing

the overall evaluation task into a set of simpler sub-
tasks, each of which is well within the judgmental
capacities of the decision maker. These procedures
typically involve the following major steps: a) list
the set of alternatives to be evaluated; b) specify the
set of attributes with respect to which each alternative
is to be evaluated; c) numerically assess the value of
each alternative with respect to each attribute; and
d) specify an arithmetic evaluation rule for determining
the overall value of each alternative. The final task,
specification of an evaluation rule, is generally accom-
plished in one of two ways. When magnitude estimation
procedures are used, each attribute is assigned a
quantitative importance factor. The overall value of each
alternative is then computed, usually using a weighted
sum or product. When indifference judgment procedures are
used, the decision maker must first select an important
and continuous base attribute against which all other
attributes can be traded off. Then trade-offs are assessed
between the base attribute and each of the other attributes.
Under the assumption that value combines additively across
dimensions, the overall value of each alternative can then
be expressed in terms of units of the base attribute.

Proponents of the decomposition approach claim the
following major advantages. First, by reducing the
information burden placed upon the decision maker, decom-
position procedures should substantially reduce the amount
of random error in the evaluation process. Second,
decomposition permits decision makers to consider a much
larger number of attributes in choosing between alternatives.
These are probably the two most important advantages of
decomposition over wholistic judgment. Others have been
cited, however. Raiffa (1969) has argued that it can
assist decision makers in the logical structuring of their
problems, and Edwards (1971) has discussed a case study
in which the explicitness of the procedures facilitated
both the communication and resolution of conflicts between
decision makers representing divergent interests. Edwards
also argued that decomposition is especially suited to
application in organizational environments, with specialists
making judgments in their own areas of expertise and decision
makers with overall responsibility specifying and weight-
ing the criteria. Finally, Yntema and Torgerson (1961)
have argued that decomposition should be particularly use-
ful in decision making contexts requiring a large number
of routine value judgments. For, once developed, a quanti-
tative evaluation function can be incorporated in a computer
algorithm, thus freeing the decision maker for intellectually
more challenging tasks.

DECOMPOSED EVALUATION: THEORY, SENSITIVITY, METHOD, AND VALIDATION

As outlined above, decomposition procedures are rather ad hoc in nature. Critics might argue that the weighted sum formulation ignores configural interactions between attributes, interactions which decision makers themselves claim to take into consideration. This objection is greatly weakened, however, by the large body of experimental literature which shows that wholistic judgments can be very well approximated by simple additive models, and that configural considerations account for but a very small proportion of variance in this evaluative process (Slovic and Lichtenstein, 1971). Nevertheless, a stronger normative justification of decomposition procedure is required.

The Theoretical Basis for Multi-Attribute Evaluation

Riskless Decision Making. A decision is said to be riskless when the decision maker is able to specify with certainty the consequences associated with each course of action. Thus, riskless decisions require that the decision maker select one from a set of outcomes. When the riskless choice assumption is appropriate, the theory of conjoint

measurement (Krantz, Luce, Suppes, and Tversky, 1971) provides a formal axiomatic basis for additive decomposed evaluation. Notationally, let $(x_{1i}, x_{2i}, \ldots, x_{ni})$ be the vector of attributes describing outcome X_i, with x_{ji} denoting the j-th attribute of outcome X_i. Further, let $X_r \overset{.}{<} X_s$ denote the relationship "outcome X_r is not preferred to outcome X_s", and let $X_r \sim X_s$ denote "the decision maker is indifferent between outcomes X_r and X_s."

The first two assumptions of the conjoint measurement formulation are fundamental to all theories of rational choice (Arrow, 1952). These are:

1) <u>Connectedness</u>: For any two outcomes X_i and X_j, either $X_i \overset{.}{<} X_j$, $X_j \overset{.}{<} X_i$ or $X_i \sim X_j$.

2) <u>Transitivity</u>: For any three outcomes, X_i, X_j, X_k, if $X_i \overset{.}{<} X_j$, and $X_j \overset{.}{<} X_k$, then $X_i \overset{.}{<} X_k$.

When both of these assumptions are satisfied, preferences are said to be <u>weakly ordered</u>.

Next we consider the two independence assumptions which are at the heart of the conjoint measurement form of additivity. When outcomes are characterized by only two attributes, the following independence assumption is required.

3a) <u>Cancellation</u>. Let x_{1i}, x_{1j}, and x_{1k} be any three states of the first attribute x_1, and let x_{2i}, x_{2j}, and x_{2k} be any three states of the second attribute x_2. If $(x_{1j}, x_{2i}) \lesssim (x_{1k}, x_{2j})$ and $(x_{1i}, x_{2j}) \lesssim (x_{1j}, x_{2k})$, then $(x_{1i}, x_{2i}) \lesssim (x_{1k}, x_{2k})$.

When outcomes are characterized by three or more attributes, each of which is sufficiently important to influence the decision maker's preferences, then the cancellation axiom is replaced by the following independence assumption.

3b) <u>Monotonicity</u>. Let $(x_1, x_2, \ldots, x_n)$ be the attribute vector describing the generic outcome X. Let Y be any subset of these attributes and let Z be the vector of remaining attributes, so that $X = (Y, Z)$. Let Y_i and Y_j be any two states of the Y attributes, and let Z_i and Z_j be any two states of the Z attributes. Then $(Y_i, Z_i) \lesssim (Y_j, Z_i)$ if and only if $(Y_i, Z_j) \lesssim (Y_j, Z_j)$.

Intuitively, preferences for states of the Y attributes are not influenced by the state in which the Z attributes are held fixed.

Although the theory of conjoint measurement involves
other technical assumptions, for practical purposes
satisfaction of weak ordering and cancellation in the two
attribute case and of weak ordering and monotonicity in
the case of three or more attribute., is necessary and
sufficient to guarantee the existence of an additive
evaluation function for riskless choice (Krantz, Luce,
Suppes, and Tversky, 1971). That is, there will exist an
additive value function V comprised of constituent functions
V_1, V_2, ..., V_n such that, for any two outcomes X_i and X_j,
$X_i \stackrel{.}{<} X_j$ if and only if $V(X_i) \leq V(X_j)$, where

$$V(X) = V_1(x_1) + V_2(x_2) + \ldots + V_n(x_n).$$

Further, because riskless choice requires only that the
decision maker rank order outcomes in terms of their
desirability, if V is a value function, then any monotone
transformation of V is also a value function.

Edwards (1971) has argued that it is difficult to
imagine circumstances under which the assumptions required
for additivity are not satisfied. Those familiar with the
economic concepts of complementary and competing goods
might object. For example, the attribute "number of
right shoes" and the attribute "number of left

shoes" clearly do not combine additively in deter-
mining the overall value of a commodity bundle of
clothing. From a practical standpoint, however, this
example need not lead to the rejection of additive
evaluation rules; for the attribute "number of pairs of
shoes" may well contribute additively to the overall value
of the commodity bundle. Rather, this example illustrates
that satisfaction of additivity depends upon an appropriate
definition of attributes. In general, Edwards' assertion
seems sound. Additive evaluation models should be appro-
priate for most riskless decisions.

Risky Decision Making. When decision making involves
uncertainty, on the other hand, the assumption required to
guarantee the existence of an additive evaluation function
is strong and intuitively unappealing. Thus, risky
decision making may frequently require non-additive evalua-
tion procedures. Formally, a decision is said to be
risky when, for each possible course of action, the
decision maker is able to specify a probability distribu-
tion over the possible consequences of that action. Let
$(A_1, A_2, \ldots, A_m)$ be the set of possible actions and let
$(X_1, X_2, \ldots, X_n)$ be the set of possible consequences of those
actions. Then for each act A_i there is an associated

probability distribution of outcomes $(p_{11}, X_1; p_{21}, X_2;$
$\ldots; p_{ni}, X_n)$. That is, given that act A_i is selected,
outcome X_1 will occur with probability p_{1i}; outcome X_2
with probability p_{2i}; and so on. Thus, in *risky* decision
making the decision maker chooses not between outcomes,
but rather between probability distributions of outcomes.

A number of strategies for making such decisions
have been proposed, but the expected utility principle has
generally come to dominate normative discussions of risky
choice (Luce and Raiffa, 1957). According to this principle
there exists a utility function U defined on outcomes
such that:

a) For any two outcomes X_i and X_j, $X_i \lesssim X_j$ if and
   only if $U(X_i) \leqq U(X_j)$.

b) For any two actions A_i and A_j, $A_i \lesssim A_j$ if
   and only if $EU(A_i) \leqq EU(A_j)$.

Here $U(X_i)$ deontes the utility of outcome X_i and $EU(A_i)$
denotes the expected utility associated with action A_i
where

$$EU(A_i) = p_{1i}U(X_1) + p_{2i}U(X_2) + \ldots + p_{ni}U(X_n).$$

Finally, the utility function U is defined on an interval
scale; that is, if U is a proper utility function, then any
positive linear transformation of U is also a proper utility
function.

The expected utility principle is not new; Bernoulli
(1738) discussed it over 200 years ago. Its status as
a normative principle was not firmly established, however,
until von Neumann and Morgenstern (1944) demonstrated
that it could be derived from a set of basic axioms of
rational choice. Since that time a number of other
axiomitizations of the expected utility principle have
appeared (Herstein and Milnor, 1953; Savage, 1954; Luce
and Raiffa, 1957; Krantz, Luce, Suppes, and Tversky, 1971).

As stated above, the expected utility principle is
neutral with respect to the description of outcomes; they
may be either single- or multi-attributed. And when out-
comes are multi-attributed, the theory is neutral with
regard to the composition rule relating each attribute of
an outcome to the overall utility of the outcome. Fishburn
(1965), however, has specified a single additional assump-
tion which, when combined with the expected utility
principle, guarantees that this function will be additive.
Central to Fishburn's proof is a relationship between finite
gambles (or discrete probability distributions over finite
sets of outcomes) which we will term marginal equivalence.
Two gambles are said to be marginally equivalent if they
give rise to identical marginal probability distributions
over the possible states of each outcome attribute. This

concept is most easily illustrated for the case of binary

attributes. Consider outcomes of the form (x_1,x_2,x_3)

where the first attribute may assume either of the states

x_1' or x_1''; the second attribute either of the states

x_2' or x_2''; and the third attribute either of the states

x_3' or x_3''. Next, consider the gambles G_1 and G_2 where

$$G_1 = \begin{cases} \text{with probability } 1/3 \text{ receive outcome } (x_1'',x_2',x_3') \\ \text{with probability } 1/3 \text{ receive outcome } (x_1',x_2'',x_3') \\ \text{with probability } 1/3 \text{ receive outcome } (x_1',x_2',x_3'') \end{cases}$$

$$G_2 = \begin{cases} \text{with probability } 2/3 \text{ receive outcome } (x_1',x_2',x_3') \\ \text{with probability } 1/3 \text{ receive outcome } (x_1'',x_2'',x_3'') \end{cases}$$

For both gambles the probabilities of receiving attribute

states x_1', x_2', and x_3' are 2/3 and the probabilities of

receiving attribute states x_1'', x_2'', and x_3'' are 1/3.

Thus, G_1 and G_2 are marginally equivalent. For though the

joint probability distributions over outcome attributes

differ for the two gambles, the marginal distributions are

the same. Fishburn has shown that a multi-attribute utility

function U defined on a finite set of outcomes can be

additive if and only if the following assumption is satisfied.

4) <u>Marginality</u>: Let G_1 and G_2 be any two
finite gambles defined on the outcome set.
If G_1 and G_2 are marginally equivalent,
then $G_1 \sim G_2$.

That is, given that the expected utility principle is
satisfied, an additive utility function U exists if and
only if the decision maker's preferences satisfy the
marginality principle. And, as the following example
illustrates, the marginality assumption is very strong, and
may in many cases fail to be satisfied. Consider the
following two gambles.

$$G_a = \begin{cases} \text{with probablity 1/2 receive \$5000 and a 1973 Volvo} \\ \text{with probability 1/2 receive \$10 and a rusty hubcap} \end{cases}$$

$$G_b = \begin{cases} \text{with probability 1/2 receive \$5000 and a rusty hubcap} \\ \text{with probability 1/2 receive \$10 and a 1973 Volvo} \end{cases}$$

Since G_a and G_b are marginally equivalent, a utility func-
tion defined on dollars and automobile components can be
additive if and only if the decision maker is indifferent
between G_a and G_b. A casual survey indicates that most
people are not; they prefer G_b which provides a sure pros-
pect of attaining a highly valued outcome. More generally,

it would appear that the marginality assumption will be violated in a wide variety of contexts and that, consequently, non-additive evaluation functions will be required for many risky decision making contexts.

Fortunately, a simple formulation of non-additive utility assessment follows quite naturally from the formal definitions of value and utility. Recall that V is said to be a value function if, for any two outcomes X_i and X_j, $X_i \overset{\cdot}{<} X_j$ if and only if $V(X_i) \leqq V(X_j)$. Recall also that U is said to be a utility function whenever the following two conditions are satisfied: a) for any two outcomes X_i and X_j, $X_i \overset{\cdot}{<} X_j$ if and only if $U(X_i) \leqq U(X_j)$; and b) for any two actions A_k and A_l, $A_k \overset{\cdot}{<} A_l$ if and only if $EU(A_k) \leqq EU(A_l)$. From these definitions it is clear that U and V, if they exist, must be monotonically related. That is, if there exist bonafide utility and value functions U and V, respectively, then there will exist a monotonic transform R such that $U(X_i) = R(V(X_i))$. Thus, given that an appropriate value function V has been assessed, determination of U requires only that R be specified. This formulation applies whether or not U is additive. Note that V may be additive while U is not. For example, let $(x_1, x_2, \ldots, x_n)$ be the vector of attributes characterizing

the generic outcome X, and let $V(X) = \Sigma V_j(x_j)$. Further, let $U(X) = \log_{10} V(X)$. Then, $U(X) = \log_{10} (\Sigma V_j(x_j))$, which is not additive in the x_j. This case of additive value but non-additive utility will arise whenever the weak ordering, monotonicity, and expected utility assumptions are satisfied, but the marginality assumption is not. Throughout this paper the approach discussed here will be termed the R(V) method of multi-attribute utility assessment.

A Sensitivity Analysis

The arguments of the previous section suggest that while additive evaluation rules will probably be appropriate for most riskless decision making contexts, non-additive rules may be required for many risky contexts. Yntema and Torgerson (1961) have argued, however, that whenever the monotonicity assumption is satisfied, additive main effects models will provide an excellent approximation to overall value, regardless of how highly interactive the true data generator. This hypothesis implies that the distinction between additive and non-additive evaluation rules is trivial from a practical standpoint. It has also been argued that additive evaluation models are very robust with respect to both the assessment of the functions

relating each attribute to overall value (Edwards, 1971)
and to the specification of weighting factors for the
attributes (O'Connor, 1972; Fischer and Peterson, 1972).
In this section numerical examples are constructed to
test each of these hypotheses.

Additive Approximations to Non-Additive Composition
Rules. The theory of conjoint measurement assures that
whenever the weak ordering and monotonicity assumptions
are satisfied, there exists an additive evaluation rule
which will preserve the ordinal properties of the decision
maker's preferences. For risky decision making, however,
an evaluation function must also reflect the interval
scale properties of the decision maker's preferences, and
satisfaction of monotonicity is not sufficient to assure
that an additive function can preserve these properties.
Yntema and Torgerson (1961) argued, however, that if
monotonicity is satisfied, then an additive approximation
will do an excellent job of preserving these interval scale
properties. They supported this argument with a simple
numerical example in which an additive approximation to a
data generator consisting only of two way multiplicative
interaction terms accounted for 94% of the variance. The
present analysis tested the Yntema and Torgerson hypothesis

in the presence of higher order interactions as well.
The conclusions drawn are relevant only to decision
making contexts in which an interval scale measure of
value or utility is required.

This analysis considered only two classes of functions
satisfying the monotonicity condition, and is, therefore,
illustrative rather than exhaustive. Functions of the
first class consisted of additive and two-way cross product
terms.

$$F_1(x_1, x_2, \ldots, x_n) = \sum_i x_i + \sum_i \sum_j x_i x_j$$
$$\substack{i \neq j}$$

These functions closely resemble those studied by Yntema
and Torgerson (1961), which included only the two-way
cross product terms. Functions in the second class con-
sisted of additive and n-way cross product terms.

$$F_2(x_1, x_2, \ldots, x_n) = \sum_i x_i + \prod_i x_i$$

To assess the ability of additive models to approxi-
mate F_1 and F_2, 1000 vectors of attribute values were
generated for alternatives described by three, six, and
nine attributes. To avoid discontinuities associated with
multiplying by zero, values with respect to each attribute

were constrained to be greater than or equal to one. Actual attribute values were randomly generated from a uniform distribution over the range 1 to 100. This data generating process was such that attributes were not correlated with one another.

Using this data set, additive main effects approximations were correlated with F_1 and F_2 for the three, six, and nine attribute alternative sets (see Table 1). In general, the additive approximations to the F_1 models are excellent, and their quality improves as the number of attributes increases from three to nine. Additive main effects models do not, however, provide good approximations to the F_2 models, and the quality of these approximations declines sharply as the number of attributes increases.

These results clearly reveal that the Yntema and Torgerson example has been overinterpreted. Additive models do not provide good approximations to highly interactive data generators satisfying the monotonicity condition. From the standpoint of applied decision theory, this demonstrates that when the marginality assumption is seriously violated, additive evaluation functions will not be acceptable for risky decision making.

TABLE 1

PRODUCT MOMENT CORRELATIONS FOR ADDITIVE
APPROXIMATIONS TO NON-ADDITIVE COMPOSITION RULES

N^a	Composition Rule	
	$\Sigma X_i + \Sigma\Sigma X_i X_j$	$\Sigma X_i + \pi X_i$
3	.964	.858
6	.985	.678
9	.990	.480

[a] Number of attributes

<u>Linear</u> <u>Additive</u> <u>Approximations</u> <u>to</u> <u>Non-Linear</u> <u>Additive</u>
<u>Models</u>. The last analysis demonstrated that evaluation
models are considerably more sensitive to the proper
specification of a composition rule than has generally
been realized. This analysis tested the sensitivity of
multi-attribute evaluation rules to the specification of
functions relating each attribute to overall worth.
Throughout this analysis, additive composition rules were
assumed to be appropriate.

Given the unlimited number of functions which might
conceivably arise, an exhaustive analysis was again unfeasible.
Instead, only four monotone functions were considered. The
first two of these were members of a family of exponential
functions frequently discussed in the utility theory
literature:

$$f_1(x) = a_1(1 - e^{-x/50})$$

$$f_2(x) = a_2(1 - e^{-x/10}).$$

Here a_1 and a_2 are scaling constants such that $f_1(0) = 0$
and $f_1(100) = 100$. Figure 1 displays plots of these two
functions.
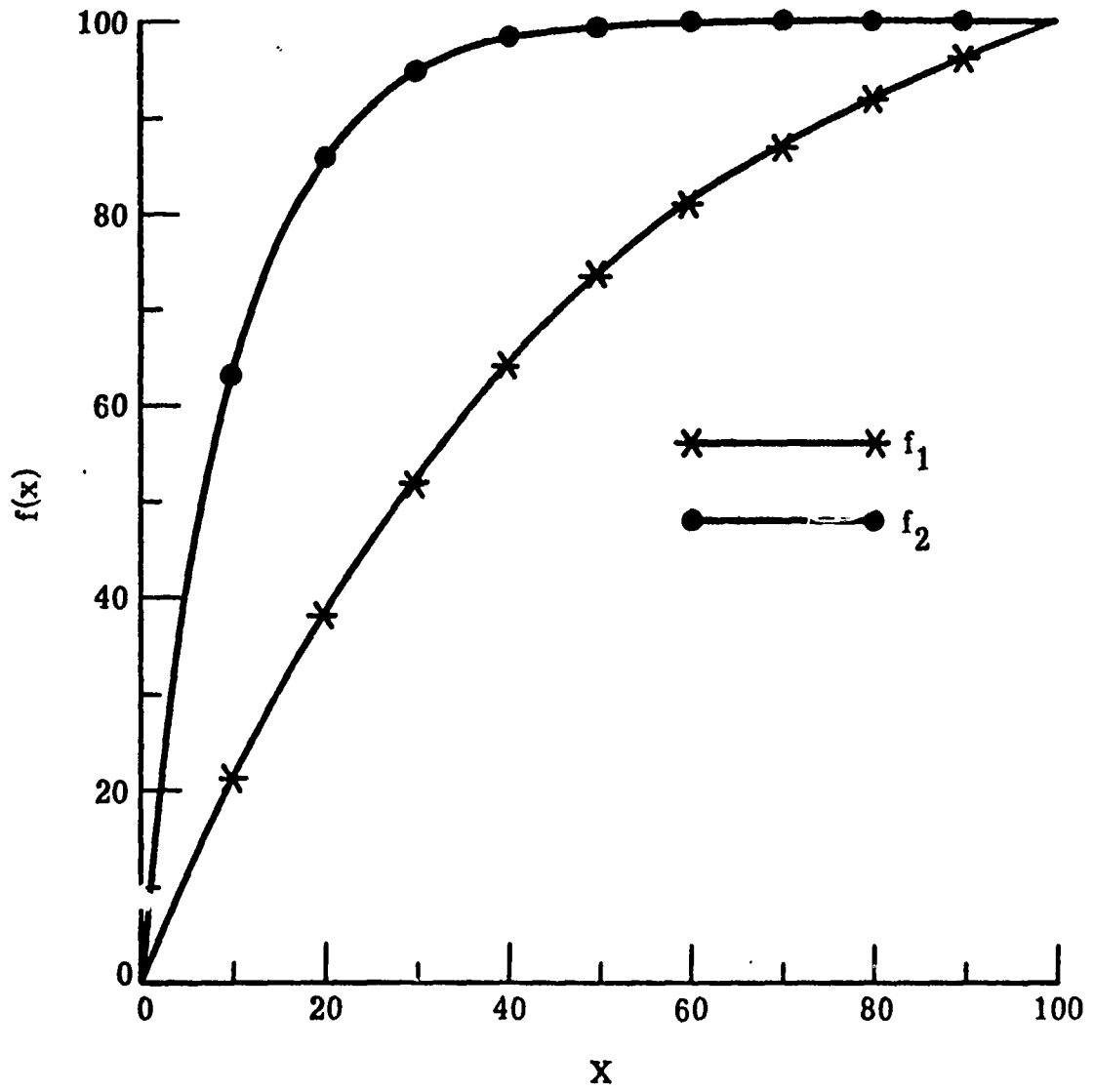
Figure 1. Plots of the functions f_1 and f_2.

The third and fourth functions studied were members of a family of power functions of a type frequently encountered in studies of psychophysical judgment:

$$f_3(x) = a_3(x - 70)^{1/3} + b_3$$

$$f_4(x) = a_4(x - 20)^{1/5} + b_4.$$

Graphs of these two functions are displayed in Figure 2. Again, a_3, b_3, a_4, and b_4 are scaling constants such that $f_1(0) = 0$ and $f_1(100) = 100$.

Using these functions, four classes of multi-attribute evaluation models were constructed:

$$F_1(x_1, x_2, \ldots, x_n) = f_1(x_1) + f_1(x_2) + \ldots + f_1(x_n)$$

$$F_2(x_1, x_2, \ldots, x_n) = f_2(x_1) + f_2(x_2) + \ldots + f_2(x_n)$$

$$F_3(x_1, x_2, \ldots, x_n) = f_3(x_1) + f_3(x_2) + \ldots + f_3(x_n)$$

$$F_4(x_1, x_2, \ldots, x_n) = f_4(x_1) + f_4(x_2) + \ldots + f_4(x_n).$$

Actual numerical examples were constructed for n = 1, 3, 6, and 9. For the single attribute case, values were computed for all integer values of $\underline{x}$ between 0 and 100. For the
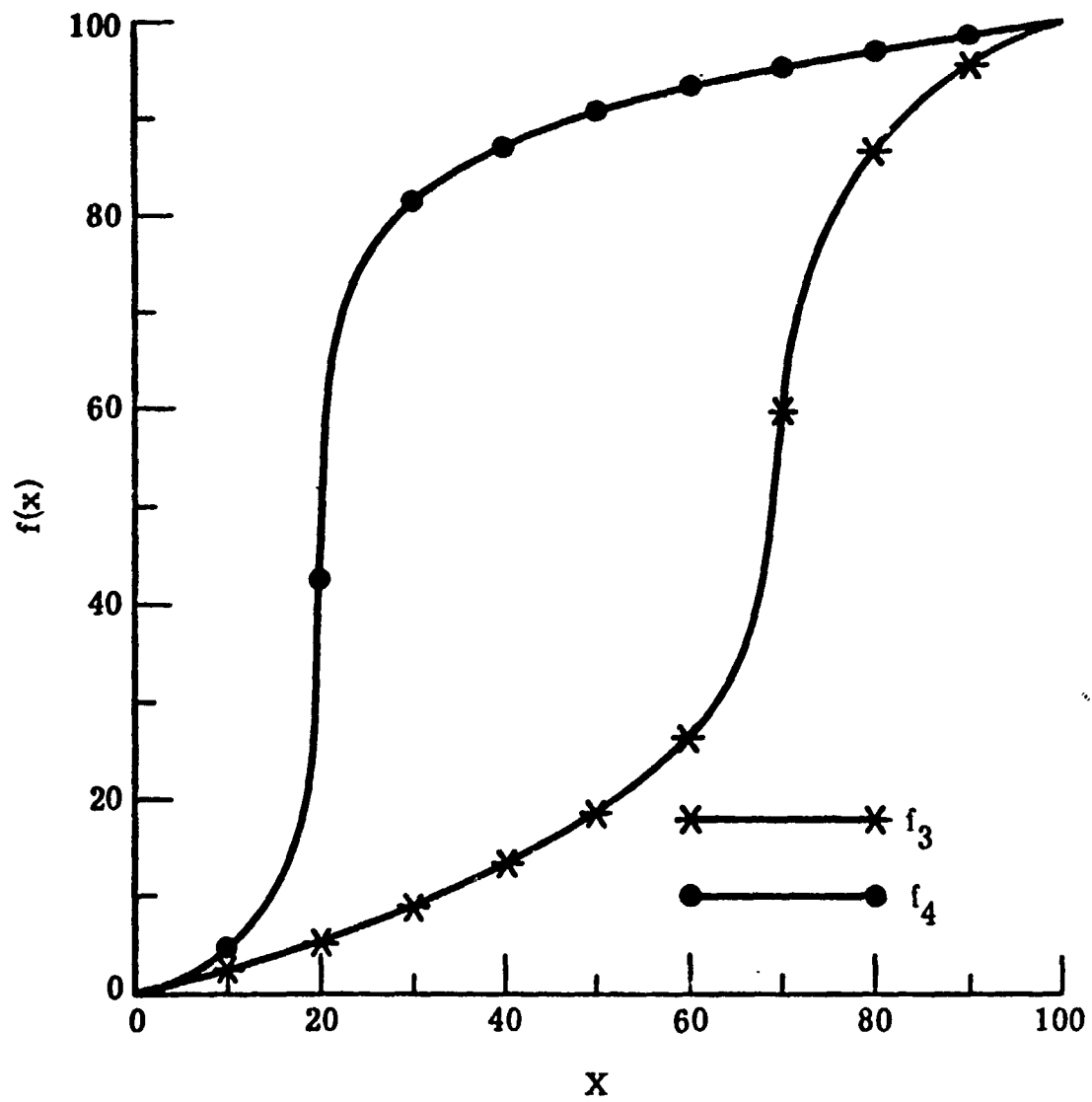
Figure 2. Plots of the functions f_3 and f_4.

multi-attribute cases, 500 three, six, and nine attribute
vectors were randomly generated. States of each attri-
bute were randomly selected from a uniform distribution
over the range 0 to 100. Again, this data generator was
such that attributes were uncorrelated. Overall values
were computed for each of these alternative vectors using
F_1, F_2, and F_4.

In order to assess the sensitivity of these multi-
attribute evaluation rules to the proper specification of
component functions, three types of approximations were
considered. For the "straight line" approximations, value
was assumed to increase linearly with x; in particular, it
was assumed that $f(x) = x$. For the "one-step" approxima-
tions, the true value for $x = 50$ was plotted. Straight line
segments were then used to connect this point, $(50, f(50))$,
with the two end points of the scale, $(0,0)$ and $(100,100)$.
For the "three-step" approximations, true values were
plotted for $x = 25$, 50, and 75. Straight line segments
were then used to connect these points, $(25, f(25))$,
$(50, f(50))$, and $(75, f(75))$, with the end points of the
scale. Separate approximations were developed for each
of the four functions studied. Additive models based on
these approximations were then used to compute overall values
for each of the randomly generated alternatives. These

approximated values were then correlated with the true values generated by F_1, F_2, F_3, and F_4. The results of this analysis are presented in Table 2.

The first important result of this analysis is that the quality of the approximations does not vary with the number of attributes describing an alternative. Overall correlations with the multi-attribute alternatives are not substantially better than the correlations of the approximations to the individual component functions. Thus, straight line approximations provide a good fit only when the component functions are highly linear (i.e., in the cases of F_1 and F_3). In all cases, however, the three-step approximations are excellent. And, except in the case of F_2, the one-step approximations are also very good.

These results suggest that while straight line approximations will not always be acceptable, it is not necessary to precisely assess the value of every state of a given attribute. Even in the case of continuous attributes, it should seldom be necessary to assess the value of more than three to five intermediately valued states. Interpolation between these points should provide an excellent approximation to the value of other states of the attribute.

TABLE 2

PRODUCT MOMENT CORRELATIONS FOR LINEAR ADDITIVE
APPROXIMATIONS TO NON-LINEAR ADDITIVE FUNCTIONS

Function	N^a	n^b	Approximation		
			Linear	1-Step	3-Step
f_1	1	101	.968	.997	1.000
	3	500	.972	.997	.999
	6	500	.971	.997	.999
	9	500	.970	.996	.999
f_2	1	101	.691	.851	.971
	3	500	.727	.863	.974
	6	500	.710	.859	.973
	9	500	.705	.851	.969
f_3	1	101	.925	.981	.981
	3	500	.924	.981	.980
	6	500	.928	.981	.980
	9	500	.927	.981	.980
f_4	1	101	.825	.936	.968
	3	500	.835	.935	.971
	6	500	.833	.934	.971
	9	500	.834	.936	.966

[a] Number of attributes.

[b] Number of observations upon which correlation is based.

Sensitivity to Weighting Parameters. This section
considers the sensitivity of evaluation rules to inaccurate
assessment of weighting parameters. Throughout it will be
assumed that an additive composition rule is appropriate
and that the component functions have been accurately
assessed.

For three, five, and nine attribute outcomes, three
different additive evaluation models were constructed.
These models differed only in the relative weight which
they assigned to each attribute. In each case, the first
of these three rules discriminated very highly between
the attributes, assigning 81 times as much weight to the
most important attribute as to the least important attri-
bute. The second rule utilized the same rank ordering
of attribute importance, but afforded much less discrimina-
tion. Here the ratio of the weights assigned to the most
and least important attributes, respectively, was only nine
to one. The third rule did not discriminate at all between
attributes; each received equal weight. Specifically, for
the three attribute alternatives the "high", "low", and "no"
discrimination models were:

$$H_3(x_1,x_2,x_3) = 81x_1 + 25x_2 + x_3$$

$$L_3(x_1,x_2,x_3) = 9x_1 + 5x_2 + x_3$$

$$N_3(x_1,x_2,x_3) = x_1 + x_2 + x_3.$$

For the five attribute outcomes the three evaluation rules were:

$$H_5(x_1, x_2, \ldots) = 81x_1 + 49x_2 + 25x_3 + 9x_4 + x_5$$

$$L_5(x_1, x_2, \ldots) = 9x_1 + 7x_2 + 5x_3 + 3x_4 + x_5$$

$$N_5(x_1, x_2, \ldots) = x_1 + x_2 + x_3 + x_4 + x_5.$$

And for the nine attribute outcomes the rules were:

$$H_9(x_1, x_2, \ldots) = 81x_1 + 64x_2 + 49x_3 + 36x_4 + 25x_5$$
$$+ 16x_6 + 9x_7 + 4x_8 + x_9$$

$$L_9(x_1, x_2, \ldots) = 9x_1 + 8x_2 + 7x_3 + 6x_4 + 5x_5$$
$$+ 4x_6 + 3x_7 + 2x_8 + x_9$$

$$N_9(x_1, x_2, \ldots) = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$
$$+ x_7 + x_8 + x_9.$$

Each of these nine evaluation rules was then used to compute the value of 1000 randomly generated attribute vectors. Again, the attributes were uncorrelated and values within each attribute were generated by a uniform distribution over the range of 0 to 100. Finally, for three,

five, and nine attributes, the high, low, and no discrimina-
tion models were correlated with one another. These
correlations, displayed in Table 3, indicate that additive
models are quite robust aga'.nst improper specification of
weighting parameters. The low discrimination models
afford excellent prediction of the high discrimination
models. In addition, the no discrimination models provide
a good approximation to the low discrimination models.
These results suggest that additive evaluation rules are
relatively insensitive to errors in the assessment of
importance weights. The low correlation between the high
discrimination and no discrimination models demonstrates,
however, that insensitivity is a relative matter and that
gross errors in the assessment of weights will lead to
seriously biased overall value assessments.

Methods for Assessing Decomposed Evaluation Functions

In the previous sections we have considered the
theoretical basis for multi-attribute evaluation and the
sensitivity of multi-attribute evaluation rules to
various types of assessment errors. This section discusses
four procedures for assessing a decomposed evaluation
function. Throughout this discussion it is assumed, first,

TABLE 3

PRODUCT MOMENT CORRELATIONS BETWEEN ADDITIVE
MODELS WITH DIFFERENT WEIGHTING FACTORS

N^a	Models[b]		
	High-Low	High-No	Low-No
3	.977	.756	.855
5	.967	.736	.865
9	.968	.768	.888

[a] Number of attributes.

[b] The High, Low, and No discrimination models
assigned relative weights of 81:1, 9:1, and
1:1 to the most and least important
attributes, respectively.

that the set of alternatives to be evaluated has been
listed, and second, that the attributes with respect
to which these alternatives are to be evaluated have been
specified.

The first of the four procedures, the <u>additive rating</u>
<u>scale</u> method, has been adapted from Edwards (1971),
Fishburn (1965), and Hoepfl and Huber (1970). It is
designed for riskless evaluation and involves four major
steps. First, for each attribute, x_i, the decision maker
specifies the most and least desirable states, denoted by
x_i^* and x_{i*}, respectively. Arbitrarily, these states may
be assigned values of 100 and 0, respectively.

Next, within each attribute, the decision maker assigns
numerical values to all intermediately valued states. For
example, consider the i-th attribute and let x_{ij} be any
state of this attribute such that $x_{i*} < x_{ij} < x_i^*$. For
each such state the decision maker assesses a value
$V_i(x_{ij})$ which reflects the value of x_{ij} relative to x_{i*}
and x_i^*. If the attribute is continuous, interpolation
between a few well chosen points will be required.

Third, weighting or scaling factors are assessed
which reflect the relative importance of each attribute.

The relative weight assigned to a given attribute should
be proportional to the change in overall value produced
by moving that attribute from its least to most valued
state, all other things being equal (Fishburn, 1965).
For convenience, these weights may be normalized to sum to
one.

Finally, overall values may be assigned to each
alternative by the additive evaluation rule

$$V(x_1, x_2, \ldots, x_n) = w_1 V_1(x_1) + w_2 V_2(x_2) + \ldots + w_n V_n(x_n),$$

where w_i is the normalized weighting factor for the i-th
attribute.

The second decomposition procedure, the _additive
trade-off_ _method_, is also designed for riskless decision
making. Essential to this procedure is the existence of
an important continuous base attribute against which all
other attributes may be traded off; for example, lives
saved or equivalent dollar value. Let x_1 denote this
attribute. Next, for each attribute, a standard state,
x_1^o, must be specified. Then, differences in value between
states of the other attributes can be traded off into units
of the base attribute x_1. For example, let x_{ij} be any
state of the i-th attribute. The decision maker is asked

to specify a state of the first attribute, x_{1k}, such that
$(x_1^o, x_2^o, \ldots, x_{1-1}^o, x_{1j}, x_{1+1}^o, \ldots, x_n^o) \sim (x_{1k}, x_2^o, \ldots, x_n^o)$.
Intuitively, x_{1k} should be such that the difference in
value between x_1^o and x_{1j} is equal to that between x_1^o
and x_{1k}. One such judgment is required for each state of
the attributes x_2, x_3, ..., x_n. When attributes are con-
tinuous, interpolation will be necessary. In this way
deviations from the standard state of each attribute may
be expressed in terms of units of the base attribute.

When overall value is assumed to be linear in the
units of the base attribute, each attribute may be assigned
an overall value by summing over attributes the base
attribute equivalent value of the deviation of each attri-
bute from its standard state. For example, if the base
attribute is measured in dollar units, then within each of
the other attributes deviations from the standard state of
the attribute will be assigned dollar equivalent values.
If the state of an attribute is preferred to its standard
state, then this dollar value will be positive. If the
state of an attribute is less desirable than its standard
state, then this dollar value will be negative. Thus,
assuming that value is linear with dollars, an overall
value can be assigned to an alternative by summing over

attributes the dollar equivalent value of the deviation of each attribute from its standard state.

When riskless value is not linear in units of the base dimension, it is necessary to assess a riskless value function V over this base attribute. This function may be assessed using the direct estimation procedure described for obtaining functions over attributes in the rating scale procedure. In this way, value differences within each of the other attributes may be expressed in terms of units of the value function V. These units may then be summed across attributes to assign overall values to multi-attribute alternatives.

The next two decomposition procedures to be discussed are designed for risky decision making. Because both procedures rely upon the indifference probability method (Luce and Raiffa, 1957; Raiffa, 1968) for assessing risky utilities, this method is considered first. It can, in principle, be applied to either single- or multi-attribute outcomes. It will later be argued, however, that the method is best suited for single-attribute assessment and that decomposition procedures should be used for multi-attribute utility assessment, particularly when the number of attributes is large.

Let $(X_1, X_2, \ldots, X_m)$ be the set of outcomes to which utilities are to be assigned. In applying the indifference probability procedure the decision maker must first specify the most and least desirable outcomes in this set; let X^* and X_* denote these outcomes, respectively. The decision maker assigns utilities to all other outcomes in the outcome set by comparing them with X^* and X_*. These comparisons take the form of hypothetical lotteries or gambles. Consider any intermediately valued outcome X_1. The decision maker assigns a utility to this outcome by specifying a probability p_1 such that he is indifferent between receiving the outcome X_1 with certainty or accepting the gamble $(p_1, X^*; (1-p_1), X_*)$. One such indifference probability must be assessed for each outcome in the set. These indifference probabilities themselves provide an appropriate utility measure, provided that the decision maker has assessed them in an expected utility maximizing fashion. To show that this is the case, note that utilities are defined only on an interval scale. Thus the extreme outcomes X^* and X_* may be arbitrarily assigned utilities of 1.0 and 0.0, respectively. Then, assuming that the indifference probabilities were assessed in an expected utility maximizing fashion, the relation $X_1 \sim (p_1, X^*; (1-p_1), X_*)$ implies that $U(X_1) = p_1 U(X^*) + (1-p_1) U(X_*)$, or $U(X_1) = p_1$.

The validity of the indifference probability assessment technique rests on the assumption that the decision maker acts as an expected utility maximizer in evaluating simple gambles. Although this is surely an idealization, studies of human gambling behavior indicate that expected utility models provide a fairly close approximation to this behavior, provided that the gambles involved are of the very simple type used in the indifference probability assessment method (Davidson, Suppes, and Siegel, 1957; Coombs, Bezembinder, and Goode, 1967; Tversky, 1967). Application of the indifference probability assessment technique to complex multi-attribute alternatives, however, seems very questionable because of the heavy information load placed upon the decision maker. To reduce this information overload problem, utility decomposition procedures have been developed. As in the case of riskless evaluation, these procedures reduce the complexity of the judgments required of the decision maker.

The first of the two utility decomposition procedures which we will discuss was developed by Raiffa (1969) and will be termed the _additive utility_ procedure. In implementing this method the decision maker begins by specifying the most and least preferred states of each attribute. These states are arbitrarily assigned within attribute

utilities of 1.0 and 0.0, respectively. That is, for
each attribute, $U_i(x_i^*) = 1.0$ and $U_i(x_{i*}) = C$. Next,
utilities must be assigned to intermediately valued states
of each attribute. For example, let x_{ij} denote the j-th
state of the i-th attribute. To assign a utility to state
x_{ij} the decision maker assesses a probability p_{ij} such
that he is indifferent between obtaining state x_{ij} with
certainty or accepting the gamble $(p_{ij}, x_i^*; (1-p_{ij}), x_{i*})$,
assuming all other attributes are held constant. One such
judgment must be elicited for each state of each attribute.
Note that in making these assessments, the decision maker
needs only to consider the relative utility of different
states of a single attribute. Again, if attributes are
continuous, interpolation will be required.

Next, the decision maker must assign importance
factors to each attribute. This is accomplished by having
the decision maker assess the utility range associated with
each attribute. Notationally, let $X^* = (x_1^*, x_2^*, \ldots, x_n^*)$
and $X_* = (x_{1*}, x_{2*}, \ldots, x_{n*})$ denote the most and least
desirable multi-attribute outcomes, respectively. And let
$(x_k^*, x_{\overline{k}*})$ denote the outcome which is characterized by the
most desirable state of attribute k and the least desirable
state of all other attributes. There will be n such out-
comes. In order to assess the utility range associated with

the k-th attribute, the decision maker specifies a probability, w_k, such that he is indifferent between accepting the outcome $(x_k^*, x_{\bar{k}*})$ with certainty or accepting the gamble $(w_k, X^*; (1-w_k), X_*)$. Arbitrarily letting $U(X^*) = 1.0$ and $U(X_*) = 0.0$, it can easily be shown that w_k is a measure of the utility range of the k-th attribute, and thus, that the w_k may be used as weighting factors in an additive utility function. This method of assigning weights is the most undesirable feature of the additive utility decomposition, for it relies upon wholistic utility assessments of $\underline{n}$ multi-attribute outcomes. When the number of attributes is large, these assessments may be subject to a substantial degree of error.

Finally, overall utilities may be assigned to each multi-attribute outcome according to the additive rule

$$U(x_1, x_2, \ldots x_n) = w_1 U_1(x_1) + w_2 U_2(x_2) + \ldots + w_n U_n(x_n).$$

These utilities can be used in the computation of the expected utility associated with each course of action.

Despite the apparent similarities between this procedure and the riskless rating method, Raiffa (1969) has shown that the two will not necessarily coincide. They should be monotonically related, so that both may be used

for riskless decisions. But they need not be linearly
related, and only the additive utility method is formally
appropriate for risky decision making.

The second utility decomposition, the R(V) method,
exploits the monotone relationship between riskless value
and risky utility. Recall that if V is an appropriate
measure of riskless value, then there exists some monotone
transform R such that $U(X) = R(V(X))$. So, given a decom-
posed value function V, U can be specified simply by
determining R. This may be accomplished in either of two
ways, depending upon the manner in which the riskless
value function has been assessed.

When trade-off procedures have been used to construct
the value function V, assessment of R will be particularly
easy. Suppose, for example, that all outcomes have been
traded off into a continuous base dimension, such as
equivalent dollar value or number of lives saved per year.
Then R can be directly obtained simply by assessing a
utility function over the continuous base attribute. In
contrast to the additive utility decomposition, this vari-
ant of the R(V) method requires no wholistic multi-attri-
bute utility assessments. Rather, the decision maker needs
only to assign single attribute wholistic utilities to
three to five states of the base attribute.

When the riskless value function has been constructed using rating scale procedures, assessment of R is somewhat more difficult. Raiffa (1969) has suggested that the decision maker directly assess the utilities of a few well chosen multi-attribute outcomes. The values of these outcomes (as indicated by the rating scale decomposition) can then be plotted on one axis, and the wholistic utilities assigned to these outcomes on the other. Utilities for outcomes having other values can then be obtained by interpolating a curve through the points for which utilities have been assessed. Like the additive utility decomposition, this variant of the R(V) method requires wholistic multi-attribute utility assessments. In many cases it will require fewer, however. For while this variant of the R(V) method will require three to five such judgments, the additive utility method requires as many such judgments as there are attributes.

Validation of Decomposed Evaluation Functions

The decomposition procedures just described are not difficult to implement. Nevertheless, it remains to be shown that the judgments required can be made in a systematic and meaningful fashion. Broadly speaking, experimenters have adopted two general approaches to this problem of validating decomposed value measures.

The first approach, external validation, requires that
it be possible to specify an objective (externally defined)
criterion against which to validate the value measure.
This strategy is most likely to be useful when value
attributes are essentially predictors of the decision
maker's overall objective. For example, price to earnings
ratios, corporate earnings growth trends, and dividend
yields are often used as measures of the value of an issue
on the stock exchange. These attributes are of value,
however, only in so far as they are predictors of the
investor's true goal, expected monetary return. Thus,
investment decisions generated by a decomposed evaluation
model could be validated against subsequent monetary return.
In principle, external validation should be applicable to
a wide variety of real world decision making contexts. In
practice, however, it has been employed only in two experi-
mental studies (Yntema and Torgerson, 1962; Lathrop and
Peters, 1969), both of which found decomposed evaluation to
be slightly, but not dramatically, superior to wholistic
judgment.

The second approach to validation was proposed by Miller,
Kaplan, and Edwards (1967, 1969) and rests on the principle
of convergent or construct validity.

> 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 pur-
> porting to measure the same abstract quantity
> should covary (Miller, Kaplan, and Edwards,
> 1967, p. 367).

In the context of decomposed evaluation, logically equiva-
lent evaluation procedures should assign similar values
to alternatives.

Most applications of this strategy have examined the
degree of within subject convergence between wholistic
and decomposed judgments. When alternatives are charac-
terized by a small number of attributes, information over-
load should not be a serious problem, so it is reasonable
to expect a high degree of consistency between the two
types of judgments. Mean within subject correlations
between wholistic and decomposed judgments have typically
ranged from the low .80s to high .90s (Pollack, 1964;
Hoepfl and Huber, 1970; Pai, Gustafson, and Kiner, 1971;
von Winterfeldt, 1971), though in one case (Huber,
Daneshgar, and Ford, 1971) correlations in the low .60s
were obtained. The poor convergence in this latter case,
however, may well have been due to noise in the wholistic
responses rather than to any weakness of the decomposition
approach (see Fischer, 1972).

In each of the above convergent validation studies,
decomposed models were validated against wholistic rank-
ings or ratings. They may also be validated against real
or hypothetical choices. Two studies have adopted this
strategy (Yntema and Klem, 1965; Huber, Daneshgar, and
Ford, 1971), and both obtained a high degree of within
subject convergence.

A third variant of the within subject convergent
validation strategy has also been considered. To the
extent that they are valid indicators of a decision maker's
preferences, two or more different decomposition models
ought to assign the same values to alternatives. This
approach is particularly attractive because it can be used
in real world decision making contexts in the absence of
a known criterion and without relying upon difficult
wholistic judgments. Eckenrode (1965) obtained a high
degree of within subject convergence across six different
procedures for assessing importance weights. Von Winterfeldt
(1971), on the other hand, obtained low convergence between
weighting procedures but high convergence between additive
models based upon these weights. This latter result
reflects the robustness of additive models against minor
errors in weights.

Finally, in some instances, between subject convergence may provide an appropriate validating device. This strategy is primarily applicable to matters of expert judgment rather than personal taste. Kennedy (1971) asked professional bankers and accountants to assign weights to attributes of loan applications, and obtained a high degree of convergence between the average weights assessed by these two groups of subjects. Eckenrode (1965) and O'Connor (1972), on the other hand, obtained only moderate correlations between weighting factors assessed by different expert judges. In addition, however, O'Connor found that overall indices based upon these divergent weights assigned very similar overall values to alternatives, another instance of the robustness of additive models against variations in weights.

In summary, a number of different approaches to the validation issue have been considered, and in all cases the experimental results have supported the decomposition approach. Nevertheless, this validation literature has two major shortcomings. First, only one study (von Winterfeldt, 1971) has dealt with risky utility assessment. Second, trade-off procedures, which may well prove the most useful tool for real world decision making, have yet to be experimentally validated.

EXPERIMENT 1

The goal of this experiment was to fill two gaps in
the validation literature by experimentally validating
the additive trade-off and additive utility decompositions.
Additive rating scale methods were also studied so that
the degree of convergence between the three approaches
could be assessed and so that the comparative advantages
and disadvantages of the three methods could be determined.

Method

Design. Subjects evaluated hypothetical compact
cars described by either three or nine attributes. Each
subject utilized six different response modes, three of
which required wholistic judgments and three decomposed
judgments. In the first wholistic response mode, sub-
jects evaluated each car on a 0 to 100 rating scale. In
the second, they compared each car with a "standard car"
which was approximately average in all respects, indicated
which they preferred, and assigned a dollar value to the
difference in worth between the two cars. In the third,
subjects assigned utilities to cars utilizing the indif-
ference probability assessment procedure. After all

wholistic judgments had been made, subjects constructed decomposition models using the additive rating scale, dollar trade-off, and risky utility methods.

In analyzing the data, two versions of the within subject convergent validation strategy were considered. First, each of the three types of decomposition models was used to predict each of the three types of wholistic judgments. Second, the degree of convergence between the three decomposition models was assessed and contrasted with that between the three types of wholistic judgments.

Subjects. Six male University of Michigan students served as subjects. They were screened for prior mathematical exposure to insure that they would feel comfortable with the task. One of the six failed to complete one portion of the experiment and was dropped from all data analyses.

Alternatives. Subjects evaluated hypothetical compact cars described by either the first three or all nine of the attributes listed in Table 4. Several of these attributes, such as fuel economy, were in principle continuous. Others, such as type of transmission, were categorical. In either case, at least three and at most five different states of each attribute were used to

TABLE 4

LIST OF ATTRIBUTES USED TO DESCRIBE
ALTERNATIVES IN EXPERIMENT 1

Name of Attribute	Units	States				
		1	2	3	4	5
1) Quality of Steering and Handling[a]	****	VG	G	A[c]	P	VP
2) Fuel Economy	Miles per Gal.	22	24	26[c]	28	30
3) Comfort: Rear Seat[a]	****	A	P[c]	VP		
4) Acceleration (0 to 60 mph)	Secs.	12	14	16[c]	18	20
5) Radio Equipment	****	AM/FM	AM	None[c]		
6) Cost of Upkeep	****	Low	A[c]	High		
7) Comfort: Front Seat[a]	****	VG	G	·A[c]	P	VP
8) Stopping Distance (from 60 mph)	Feet	125	135	145[c]	155	165
9) Transmission[b]	****	4-S[c]	3-S	Auto		

[a] VG, G, A, P, and VP represent ratings of Very Good, Good, Average, Poor, and Very Poor, respectively.

[b] 4-S, 3-S, and Auto represent four speed stick shift, three speed stick shift, and automatic transmission, respectively.

[c] "Standard state" for the attribute.

construct alternatives. Table 4 lists the states
considered for each of the nine attributes.

For the three attribute alternatives, a set of 12
hypothetical cars was constructed. States of the three
attributes were randomly determined subject to the
following constraints. For the five state attributes, each
state was selected at least two times and at most three;
for the three state attributes, each was chosen four times.
A similar procedure was used to construct a set of 18 cars
described by nine attributes. For five state attributes,
each state was chosen at least three times and at most
four; for the three state attributes, each was chosen six
times.

Procedure. Subjects were first familiarized with the
nine attributes. They were instructed to regard performance
and quality indicators as reliable measures collected by
an independent automotive testing service. To provide a
frame of reference for the quantitative measures of accelera-
tion, fuel economy, and braking, subjects were provided with
actual test data for the 1971 model Volkswagen "Super
Beetle," Plymouth Satellite, and Ford Boss 351 Mustang.
Next, subjects specified the most and least desirable com-
binations of attribute states for both the three and nine
attribute alternatives. These were used as a frame of
reference throughout the experiment.

Subjects next made six sets of wholistic judgments. For each of the three wholistic response modes, subjects evaluated two sets of stimuli; one in which cars were described by three attributes and one in which they were described by nine. The same two sets of stimuli were evaluated in each of the three wholistic response modes. Subjects were randomly assigned to each of the six possible orders of the three response modes. But within a given response mode, the three attribute alternatives were always presented first. This within response mode presentation order was adopted to facilitate task learning on the part of the subjects.

Elicitation of the wholistic judgments proceeded as follows. Subjects were handed a booklet of either 12 or 18 pages for the three and nine attribute alternative sets, respectively. Each page of a booklet listed the attributes describing a particular alternative. On the bottom of each page was the response device.

For the rating scale judgments, subjects indicated their response by making a slash through a 100 millimeter line which was divided into 10 equal segments, and labeled at 10 unit intervals from 0 to 100. Subjects were instructed that the previously specified most and least desirable cars could be arbitrarily assigned values of 100

and 0 on this scale, and asked to evaluate other cars relative to these two extremes.

In the wholistic dollar trade-off response mode, subjects were first shown the list of attributes characterizing the "standard car" for both three and nine attributes (see Table 4). These cars were selected so as to be approximately average in all respects. At the bottom of each page was the question "Preferred to standard car?", to which the subject responded "yes" if he preferred the car described on the page to the standard car, and "no" if he preferred the standard car. Below this the subject indicated the dollar difference in value between the two cars.

Before assessing risky utilities, subjects were introduced to the indifference probability assessment procedure. Subjects first assigned utilities to simple gambles involving only monetary outcomes. They did not actually play these gambles. After they were familiar with the procedure, subjects used it to assign wholistic utilities to the hypothetical cars. Throughout, subjects had at their disposal a table for converting from probabilities to odds levels, but always gave their responses in probability form.

Subjects were not introduced to the idea of decomposition until after all wholistic judgments had been made. Decomposed models were constructed in the order in which subjects had made the corresponding wholistic judgments.

In constructing rating scale models, subjects first assigned values to the various states of each attribute. Using a 100 millimeter scale, they located the most and least desirable states of the attribute in question at the 0 and 100 points of the scale, respectively. They then indicated the relative values of other states of the attribute by making slashes through the scale. Importance weights were elicited in a similar fashion. First subjects ranked attributes in order of their importance. They next arbitrarily assigned an importance of 100 to the most important attribute. Then, again using a 100 millimeter scale, they assigned relative importance weights to each of the other attributes. Subjects were told to view these importance assessments as percentages. For example, a rating of 50 would indicate that an attribute was 50% as important as the most important attribute. Finally, these importance weights were normalized to sum to one.

Additive trade-off decomposition models were constructed using a slightly modified version of the procedure described earlier. Within each attribute, subjects assigned dollar

equivalent values to deviations from the standard state of each attribute. For example, let x_i^o be the standard state of the i-th attribute, and let x_{ij} be any other state of that attribute. Subjects were first asked to indicate which of these two states they preferred. They were then asked to assign a dollar value to the difference in worth between these two states. These dollar differences were then assumed to combine additively across attributes.

In constructing additive utility models, subjects used the indifference probability assessment procedure to assign within attribute utilities to the states of each attribute. In making these judgments, they compared intermediately valued states of each attribute with the most and least valued states of that attribute, assuming all other attributes to be held constant. Next, subjects assessed wholistic utilities from which importance weights were derived. For both three and nine attributes, subjects assessed multi-attribute utilities for all alternatives for which one attribute was in its most valued state and all other attributes in their least valued states. As noted earlier, such assessments provide a measure of the utility range associated with each attribute and thus provide an appropriate measure of importance. In principle, these

importance assessments should have summed to one. In practice, however, they typically did not, and so were normalized to sum to one.

During all stages of the above procedure, subjects were run individually and under close supervision.

Results

Convergence between Wholistic and Decomposed Responses. As noted earlier, the formal definitions of value and utility require that any two such measures be monotonically related. So if subjects were completely consistent in making all of their wholistic and decomposed assessments, and if their wholistic assessments satisfied the monotonicity condition, then each of the three decomposition models should perfectly predict the rank ordering of alternatives generated by each of the three wholistic response modes. To test this proposition, rank order correlations were computed (see Table 5). For the three attribute alternatives the obtained correlations were in general very high. Similar results were obtained for the nine attribute alternatives, though these correlations were generally somewhat lower. Here, however, the risky utility decomposition consistently provided poorer predictions of the rank orders generated by all three wholistic response modes.

TABLE 5

RANK ORDER CORRELATIONS (RHO) BETWEEN WHOLISTIC AND DECOMPOSED ASSESSMENTS: EXPERIMENT 1

N[a]	Wholistic Mode	Decomposition	Subject 1	2	3	4	5	Median
3	Rating Scale	Rating Scale	1.00	.85	.94	.90	.97	.94
		$ Trade-off	.98	.94	.94	.93	.99	.94
		Utility	.85	.97	.95	.87	.99	.95
3	$ Trade-off	Rating Scale	.98	.73	.95	.96	.90	.95
		$ Trade-off	1.00	.86	.95	.89	.93	.93
		Utility	.86	.87	.99	.81	.93	.87
3	Utility	Rating Scale	.96	.75	.94	.98	.98	.96
		$ Trade-off	.97	.89	.94	.96	.95	.95
		Utility	.93	.99	.94	.89	.94	.94
9	Rating Scale	Rating Scale	.95	.90	.83	.91	.86	.90
		$ Trade-off	.87	.90	.85	.93	.89	.89
		Utility	.78	.78	.80	.83	.91	.80
9	$ Trade-off	Rating Scale	.96	.82	.94	.97	.89	.94
		$ Trade-off	.93	.85	.95	.99	.85	.93
		Utility	.77	.88	.95	.93	.85	.88
9	Utility	Rating Scale	.94	.75	.80	.91	.91	.91
		$ Trade-off	.91	.77	.87	.93	.94	.91
		Utility	.77	.76	.76	.84	.95	.77

[a] Number of attributes.

The utility theory axioms also require that any two interval scale measures of value or utility be linearly related. Thus, assuming risky utilities to be additive, each of the three decompositions should be equally predictive of the interval scale properties of each of the three types of wholistic responses. To test this hypothesis, product moment correlations between wholistic and decomposed judgments were computed (see Table 6). Data for the three attribute alternatives generally supported this hypothesis, though there was some indication that the risky utility decomposition afforded poorer prediction of the riskless wholistic assessments than did the riskless decomposition models. Similar results were obtained for the nine attribute alternatives, though here it was more evident that the risky utility decomposition afforded poorer prediction of the riskless wholistic assessments.

To provide a more sensitive measure of interval scale convergence, mean absolute deviations (MABS) between wholistic and decomposed assessments were also computed. Because the various sets of responses were measured on different scales, these differences could not be directly computed. To make all sets of responses comparable, each subject's data was transformed as follows. Within each response mode and stimulus set all judgments were linearly

TABLE 6

PRODUCT MOMENT CORRELATIONS BETWEEN WHOLISTIC AND

DECOMPOSED ASSESSMENTS: EXPERIMENT 1

N[a]	Wholistic Mode	Decomposition	Subject					Median
			1	2	3	4	5	
3	Rating Scale	Rating Scale	.99	.92	.95	.96	.94	.95
		$ Trade-off	.97	.96	.95	.96	.96	.96
		Utility	.80	.92	.95	.93	.97	.93
3	$ Trade-off	Rating Scale	.99	.86	.96	.96	.92	.96
		$ Trade-off	1.00	.92	.97	.92	.92	.92
		Utility	.85	.90	.98	.87	.94	.90
3	Utility	Rating Scale	.81	.80	.95	.92	.95	.92
		$ Trade-off	.85	.86	.95	.92	.95	.92
		Utility	.91	.94	.94	.93	.93	.93
9	Rating Scale	Rating Scale	.97	.96	.93	.91	.86	.93
		$ Trade-off	.93	.96	.94	.94	.89	.94
		Utility	.80	.88	.91	.86	.90	.88
9	$ Trade-off	Rating Scale	.98	.88	.96	.96	.92	.96
		$ Trade-off	.98	.89	.97	1.00	.91	.97
		Utility	.85	.90	.96	.92	.90	.90
9	Utility	Rating Scale	.80	.84	.91	.87	.89	.87
		$ Trade-off	.80	.85	.95	.84	.90	.85
		Utility	.82	.85	.89	.84	.92	.85

a Number of attributes.

transformed to range between 0 and 100. Such transfor-
mations were appropriate because the value and utility
functions in question were defined only on an interval
scale. Using this transformed data set, within subject
MABS scores were computed (see Table 7). This analysis
revealed a clear departure from the utility theory assump-
tion that preference is risk variant. For both three and
nine attributes, the utility decomposition afforded better
prediction of the risky wholistic assessments and the
riskless decompositions afforded better predictions of the
riskless wholistic assessments. This result is reminiscent
of Tversky's (1967) finding that an interval scale measure
of subjective worth varies depending upon whether or not it
is based upon risky or riskless judgments. In addition,
however, the MABS analysis of the nine attribute alternatives
generated results which seem to conflict with those of the
previously discussed rank order correlation analysis. For,
using Spearman's rho as a measure of convergence, the
additive utility decomposition models provided the poorest
predictions of all three sets of nine attribute wholistic
judgments, including those wholistic judgments which were
assessed using the indifference probability utility assess-
ment procedure. Yet using MABS as a measure of convergence,
the additive utility decomposition models provided the best

TABLE 7

MEAN ABSOLUTE DEVIATIONS OF DECOMPOSED FROM
WHOLISTIC ASSESSMENTS: EXPERIMENT 1

N[a]	Wholistic Mode	Decomposition	Subject 1	2	3	4	5	Median
3	Rating Scale	Rating Scale	5	8	7	7	8	7
		$ Trade-off	7	7	7	7	9	7
		Utility	21	13	7	11	7	11
3	$ Trade-off	Rating Scale	3	13	6	9	11	9
		$ Trade-off	1	11	6	7	9	7
		Utility	17	11	6	15	10	11
3	Utility	Rating Scale	25	17	8	13	9	13
		$ Trade-off	24	15	8	15	10	15
		Utility	11	12	8	9	10	10
9	Rating Scale	Rating Scale	6	7	8	12	13	8
		$ Trade-off	8	6	8	10	13	8
		Utility	17	12	9	15	10	12
9	$ Trade-off	Rating Scale	4	10	6	8	8	8
		$ Trade-off	2	10	6	2	8	6
		Utility	16	10	6	11	11	11
9	Utility	Rating Scale	31	14	10	17	12	14
		$ Trade-off	32	14	7	19	14	14
		Utility	19	12	10	16	9	12

[a] Number of attributes

predictions of the nine attribute wholistic utility
assessments. Apparently the additive utility decomposition
procedure was subject to a considerable degree of random
error which caused it to fail to correctly order alter-
natives which were fairly similar in overall utility. In
addition, however, the additive utility decomposition
procedure did capture some of the subjects' attitude toward
risk and was thus better able to predict the interval
scale properties of the wholistic utility assessments.

Convergence between Decomposition Models. To the
extent that they are valid, two or more decomposed
evaluation models should produce the same rank ordering
over a given set of alternatives. Rank order correlations
were computed to test this proposition (see Table 8). In
general, the three decompositions did produce similar rank
orderings. The degree of ordinal convergence between the
decomposition methods was not, however, greater than that
between the various sets of wholistic responses (see Table
9). In fact, convergence between the additive utility
decomposition and the other two decompositions was poorer
than that between the wholistic utility assessments and
the other two types of wholistic assessments. This result
further supports the earlier conclusion that the additive
utility decomposition procedure is subject to a considerable
degree of error.

TABLE 8

CONVERGENCE BETWEEN DECOMPOSED EVALUATION

MODELS: EXPERIMENT 1

Measure[a]	N[b]	Models[c]	Subject					
			1	2	3	4	5	Median
Rho	3	RS-TO	.98	.93	1.00	.97	.97	.97
		RS-UF	.84	.77	.95	.89	.97	.89
		TO-UF	.85	.90	.95	.92	.99	.92
Rho	9	RS-TO	.95	.95	.95	.96	.94	.95
		RS-UF	.77	.79	.99	.97	.94	.94
		TO-UF	.85	.88	.94	.92	.93	.92
PM	3	RS-TO	.99	.98	1.00	.98	.97	.98
		RS-UF	.82	.79	.99	.94	.99	.94
		TO-UF	.85	.88	.99	.94	.97	.94
PM	9	RS-TO	.98	.99	.97	.95	.97	.97
		RS-UF	.84	.88	.99	.97	.97	.97
		TO-UF	.85	.92	.97	.91	.97	.92
MABS	3	RS-TO	4	6	0	4	6	4
		RS-UF	18	17	3	9	3	9
		TO-UF	17	13	3	10	6	10
MABS	9	RS-TO	4	4	6	8	6	6
		RS-UF	15	11	2	6	5	6
		TO-UF	15	9	6	11	8	9

[a]MABS, PM, and Rho refer to mean absolute deviations, product moment correlation, and Spearman's Rho rank order correlation.

[b]Number of attributes.

[c]RS, TO, and UF denote the additive rating scale, additive trade-off, and additive utility decompositions, respectively.

TABLE 9

CONVERGENCE BETWEEN WHOLISTIC RESPONSE
MODES: EXPERIMENT 1

Measure[a]	N[b]	Models[c]	Subject					
			1	2	3	4	5	Median
Rho	3	RS-TO	.98	.81	.96	.81	.92	.92
		RS-UF	.96	.96	1.00	.87	.95	.96
		TO-UF	.97	.86	.95	.96	.84	.95
Rho	9	RS-TO	.92	.87	.85	.95	.91	.91
		RS-UF	.91	.89	.97	.93	.96	.93
		TO-UF	.90	.83	.83	.93	.88	.88
PM	3	RS-TO	.98	.87	.96	.89	.92	.92
		RS-UF	.79	.91	1.00	.89	.94	.91
		TO-UF	.84	.80	.96	.86	.82	.84
PM	9	RS-TO	.96	.90	.96	.95	.87	.95
		RS-UF	.75	.87	.98	.81	.95	.87
		TO-UF	.78	.80	.95	.84	.84	.84
MABS	3	RS-TO	6	14	6	10	14	10
		RS-UF	29	16	3	16	9	16
		TO-UF	25	22	7	22	18	22
MABS	9	RS-TO	6	11	8	9	12	9
		RS-UF	32	15	6	19	7	15
		TO-UF	33	15	7	19	15	15

[a]MABS, PM, and Rho refer to mean absolute deviations, product
moment correlation, and Spearman's Rho rank order correlation.

[b]Number of attributes.

[c]RS, TO, and UF denote the wholistic rating scale, trade-off,
and utility response modes, respectively.

Consideration of the interval scale properties of the three decompositions also revealed a high degree of convergence. Both product moment correlation and MARS analyses indicated that the degree of interval scale convergence between the three decompositions was consistently and substantially greater than that between the corresponding wholistic response modes (see Tables 8 and 9).

Discussion

The generally high degree of convergence between not only the decomposed models themselves, but also between the wholistic and decomposed judgments strongly suggests that all three decomposition procedures can provide a meaningful measure of the decision maker's preferences. The relatively low rank order correlations between the additive utility decomposition and the other nine attribute wholistic and decomposed judgments suggests, however, that the additive utility decomposition is less reliable than the other two decomposition methods.

Experiment 1 also produced two other noteworthy results. First, the degree of interval scale convergence between the three decomposition procedures was consistently greater than that between the corresponding sets of wholistic judgments. This finding supports the contention that decomposition can improve upon wholistic evaluation by

reducing the amount of random error in the evaluation process.

The final noteworthy result was that the degree of convergence between the decomposition models and the risky wholistic assessments was less than that between the decomposition models and the riskless wholistic assessments. Although a number of explanations of this finding are possible, two seem most salient. First, the risky wholistic assessments may simply be subject to more random error, and thus, are inherently less predictable than the riskless wholistic assessments. The difficulty of the judgments required in the risky wholistic response mode makes this explanation quite plausible. A second possible explanation is that the wholistic preferences for risky alternatives were systematically non-additive, and thus, that additive decomposition models provided an imperfect approximation to the subjects' true preferences. Confirmation of this hypothesis would suggest that non-additive decomposition procedures are required for risky decision making. For there are certainly no compelling normative grounds for additive evaluation under risk. The marginality assumption which is required for additivity seems very questionable from a normative standpoint, and it is easy to construct

instances in which this assumption is and clearly should be violated. Thus, if wholistic preferences under risk exhibit evidence of substantial non-additivity, it seems reasonable to conclude that the decision maker's preferences really are non-additive, and that any decomposition procedure devised to assist him in making decisions should be capable of reflecting this non-additivity.

EXPERIMENT 2

The primary objectives of Experiment 2 were to
determine first, whether wholistic risky utility assess-
ments are substantially non-additive, and second, whether
the R(V) utility decomposition provides a better measure
of risky utility than does the additive utility decompo-
sition studied in Experiment 1. These two objectives are
closely related. For, if wholistic preferences are
systematically non-additive under risk, then a non-additive
utility decomposition procedure, like the R(V) method,
seems called for. But even if wholistic preferences under
risk are additive, the R(V) method might prove superior to
the additive utility decomposition. Experiment 1 suggested
that the additive utility decomposition might be subject
to a substantial degree of error. The R(V) method con-
siderably simplifies the evaluation process by permitting
the decision maker to separate the consideration of value
trade-offs between attributes from consideration of his
attitude toward risk. Because of its relatively greater
simplicity, the R(V) method might provide a less error
prone measure of risky preference than the additive utility
decomposition.

Method

Design. Because wholistic utility assessments were to be used to evaluate the relative desirability of additive and non-additive utility assessment procedures, a large amount of random error in these wholistic utility judgments would have posed a serious problem. To avoid this possibility, only three attribute alternatives were considered.

In the first stage of Experiment 2 subjects wholistically evaluated hypothetical job offers using both the riskless rating scale and risky utility response modes. Because the analysis of variance was to be used to provide a formal test of the additivity assumption, subjects evaluated all possible combinations of attribute states. Half of the subjects assessed the riskless ratings first, and half the risky utilities first. After all wholistic judgments had been made, subjects constructed additive rating scale, additive utility, and R(V) utility decomposition models.

Subjects. Ten male students from the University of Michigan Schools of Engineering and Business Administration served as subjects. All were either seniors or graduate students and had, as a consequence, given serious

consideration to the problem of obtaining a job following their graduation.

Alternatives. The alternatives to be evaluated were hypothetical job offers described by three attributes: annual salary, city of employment, and type of work. Each of these attributes assumed three states, yielding 27 possible job offers.

Particular states of the three attributes were specified as follows. The experimenter arbitrarily established three salary levels -- $14000, $11000, and $9000 -- and three cities -- Boston, Cleveland, and Tulsa. But given the rather divergent backgrounds and interests of the subjects, it was not possible to arbitrarily specify three types of work which would be applicable to all subjects. Instead, the experimenter asked each subject to describe three types of work, the first of which he would view as "very good" given his interests and goals, the second of which was less desirable than the first, but nonetheless "good", and the third of which was barely acceptable or "fair".

As in Experiment 1, alternatives for the wholistic judgments were presented in booklet form. Each page of a booklet listed the three attributes describing a particular job offer. Subjects evaluated each of the 27 possible job

descriptions once in each response mode. In addition, to provide a measure of wholistic reliability, subjects re-evaluated five of the alternatives for a second time. For each wholistic response mode, two booklets of wholistic alternatives were constructed. Each booklet contained 16 job descriptions. Five of these were reliability items which occurred in both booklets. The remaining 22 alternatives were randomly divided between the two booklets. The order of booklet presentation was randomized across subjects.

Procedure. After becoming familiar with the nature of their evaluation task, subjects created the descriptions of the "fair", "good", and "very good" types of work. Next they specified the most and least desirable of the 27 possible job offers. Then wholistic assessments were elicited. As in Experiment 1, subjects indicated their riskless ratings by making a slash through a 100 millimeter scale which was divided into ten equal intervals and numbered from 0 to 100.

Elicitation of the risky wholistic judgments differed from Experiment 1, however, in that a plausible scenario could be constructed. The instructions given to subjects can be paraphrased as follows.

Suppose that the two jobs which you previously
stated were the most (J*) and least (J*) attrac-
tive of the possible job offers are two of
only three job offers which you have, and
suppose that the intermediately valued offer
(J') described on the page of the booklet before
you is the third. You have been told that J*
is yours for the asking, and you may wait several
months if you wish before giving your decision.
The J' offer, on the other hand, requires an
immediate response. You can have it now if you
wish but if you wait the offer will no longer be
available. Unfortunately, while you have some
chance of receiving the J* offer, you will not
know for certain until several weeks from now,
and you cannot wait that long before stating
whether you will accept or reject J'.

This leaves you with only two alternative courses
of action. First, you can accept J', thus for-
feiting any chance to J* but assuring that you
will not have to accept J*. Or second, you can
reject J'. In this case, if you are fortunate,
you will receive the offer J*, but if not, you
will have to accept J*. Clearly, your decision
in this matter will depend upon how likely you are
to receive the J* offer. Your task in this por-
tion of the experiment is to specify, for each
offer J', a probability p' of receiving the J*
offer such that you would be indifferent between
accepting and rejecting J'.

The actual instructions were stated less formally, and
subjects found this scenario quite plausible. Some had,
in fact, experienced such a situation. This plausibility
further strengthened the case for using these wholistic
utility assessments as a criterion for validating the
decomposed utility models.

After completing the wholistic phase of the experiment, subjects constructed decomposed evaluation models. For the rating scale and additive utility decompositions they followed the procedures utilized in Experiment 1. To obtain weighting factors for the additive utility decomposition, subjects reassessed the required wholistic multi-attribute utilities. Thus, the additive utility decompositions were independent of the set of wholistic utility judgments which they were used to predict.

Construction of the R(V) utility decompositions required no additional judgments. The rating scale models provided a measure of riskless value. To determine the transform R, the experimenter utilized the three wholistic multi-attribute utility assessments from which the weighting factors of the additive utility decompositions had been determined. These three utilities were plotted on one axis, and the corresponding values assigned by the rating scale decompositions on the other. Between these three points and the two extreme outcomes, J^* and J_*, the transform R was approximated by straight line segments joining adjacent points for which utilities had been assessed. For one subject (#5) the three wholistic utility assessments were not monotonically related to the values assigned by the rating scale decomposition. In this case, the transform R was based on the two utility assessments which were nearest to the 0 and 100 points of the utility scale.

As in Experiment 1, subjects performed all tasks individually and under close supervision.

Results

Riskless Ratings. The present design permitted a direct test of the generally accepted assumption that wholistic riskless preferences are additive. The analysis of variance summarized in Table 10 revealed that despite a significant ($p \leq .01$) City by Salary interaction, additive main effects accounted for 99.2% of the fixed effects sums of squares. So, for practical purposes, the wholistic ratings were essentially additive.

The next four analyses to be discussed were all designed to assess the degree of convergence between wholistic and decomposed riskless judgments. Because the rating scale decomposition was to be used as the basis for the R(V) decomposition, it was especially important that the rating scale models provide an accurate measure of riskless preference.

The first two measures of convergence, the rank order and product moment correlation coefficients, both indicated a very high degree of convergence between the wholistic and decomposed ratings (see Table 11). The

TABLE 10

ANALYSIS OF VARIANCE FOR WHOLISTIC
RATINGS: EXPERIMENT 2

Factor	df	F-ratio	% of Total Fixed Effects Sums of Squares
City (C)	2,18	15.12**	19.6
Salary (S)	2,18	351.80**	13.1
Type of Work (W)	2,18	45.40**	66.5
C X S	4,36	4.68**	.3
C X W	4,36	2.07	.3
S X W	4,36	1.20	.1
C X S X W	8,72	1.50	.1

** $p \leq .01$

TABLE 11

CONVERGENCE OF RATING SCALE DECOMPOSITION MODELS

WITH WHOLISTIC RATINGS: EXPERIMENT 2

Measure[a]	Subject										Median
	1	2	3	4	5	6	7	8	9	10	
Rho	.94	.97	.99	.88	.95	.97	.93	.99	1.00	1.00	.970
PM	.94	.97	.99	.87	.95	.96	.89	.99	1.00	.97	.965
MABS	8	5	4	11	7	7	14	3	2	7	7

[a] MABS, PM, and Rho refer to mean absolute deviations, product moment correlations, and Spearman's Rho rank order correlations.

third measure, the mean within subject absolute deviation
(MABS) of wholistic from decomposed ratings, also
indicated a high degree of convergence (see Table 11).
The median MABS across subjects was only 7 (on a scale of
100), a very respectable result in view of the fact that
the mean MABS between wholistic reliability items was 6.

In the final and most sensitive test of convergence,
the cumulative frequency distribution of the absolute
deviations of wholistic from decomposed judgments was
plotted and compared with the cumulative frequency distri-
bution of deviations between reliability items. The two
distributions, plotted in Figure 3, were very similar.
Only for errors of one unit or less was the wholistic
reliability substantially greater than the degree of con-
vergence between wholistic and decomposed ratings. This
result suggests that the decomposition models accounted
for virtually all systematic variance in the wholistic
ratings, and thus, that the additive rating scale models
provided an accurate measure of riskless preference.

Risky Utilities. The hypothesis that wholistic
utility assessments are additive was tested by analysis
of variance (see Table 12). This analysis indicated that,
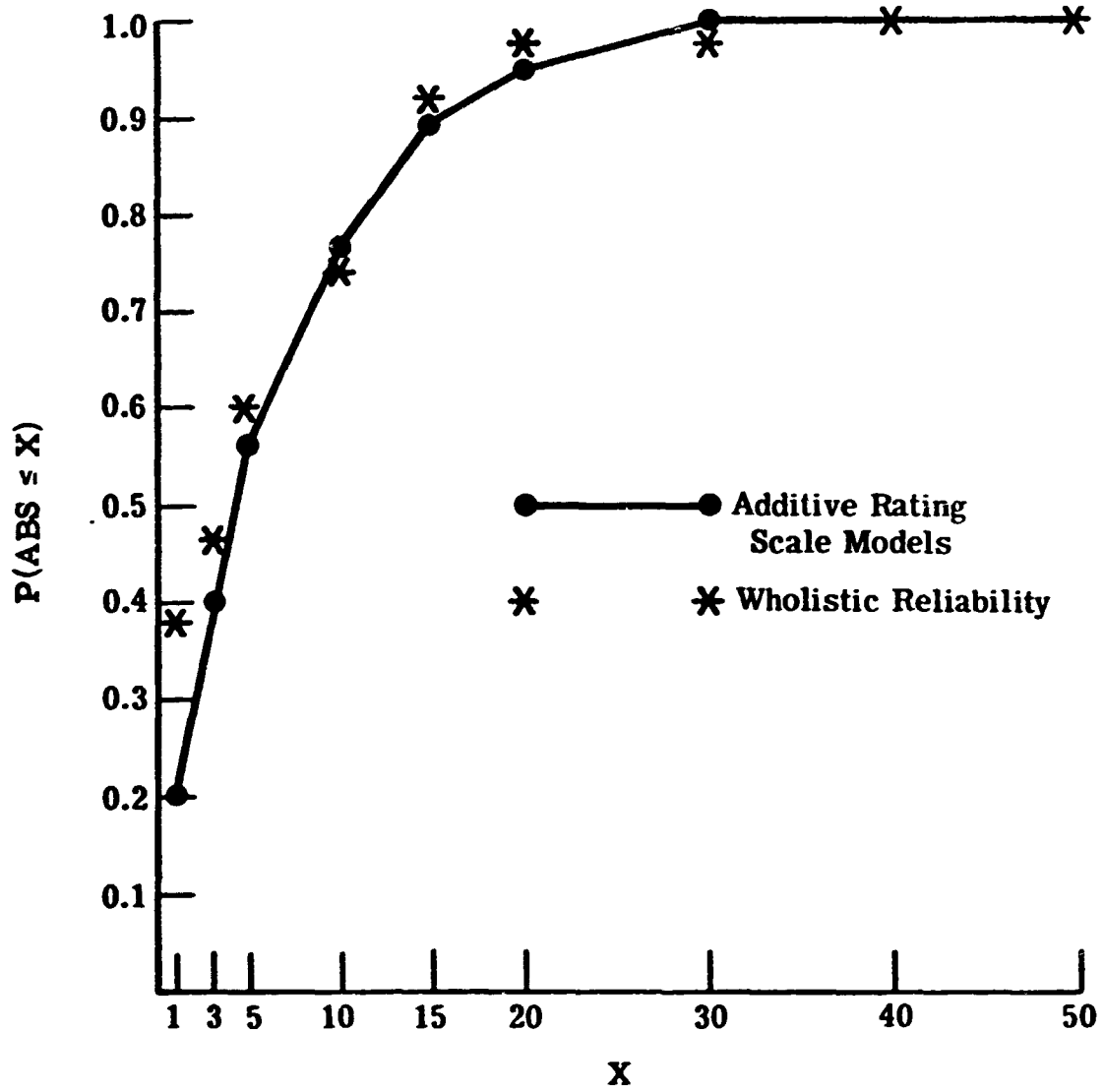despite a significant (p $\leq$ .05) City by Type of Work

Figure 3. Cumulative distributions of absolute
deviations from wholistic ratings.

TABLE 12

ANALYSIS OF VARIANCE FOR WHOLISTIC RISKY
UTILITY ASSESSMENTS: EXPERIMENT 2

Factor	df	F-ratio	% of Total Fixed Effects Sums of Squares
City (C)	2,18	23.95**	23.7
Salary (S)	2,18	492.00**	17.5
Type of Work (W)	2,18	40.10**	57.6
C X S	4,36	2.22	.1
C X W	4,36	3.10*	.9
S X W	4,36	.78	.2
C X S X W	8,72	.32	.1

** p ≦ .01

* p ≦ .05

interaction, these wholistic utility assessments were also essentially additive. Here additive main effects accounted for 98.8% of the fixed effects sums of squares.

As in the case of the riskless ratings, four measures of convergence between wholistic and decomposed judgments were obtained. In each of these analyses all three decomposition models were used to predict the wholistic utility assessments. Table 13 summarizes the rank order correlations, product moment correlations, and MABS scores. For all measures all three decomposition models yielded excellent predictions of the wholistic utility assessments. The median MABS scores for the rating scale, additive utility, and R(V) utility decompositions were 11, 9, and 9, respectively, all excellent in view of the fact that a mean MABS of 9 was obtained for the wholistic reliability judgments.

The cumulative frequency distributions of absolute deviations did, however, discriminate to some degree between the three decomposition models (see Figure 4). Examination of these distributions reveals that the R(V) and additive utility decompositions consistently dominated the rating scale models. The differences involved, however, were very small. These distributions also reveal that the

TABLE 13

CONVERGENCE OF THE THREE DECOMPOSITION MODELS WITH
THE WHOLISTIC UTILITY ASSESSMENTS: EXPERIMENT 2

Measure[a]	Model[b]	Subject										Median
		1	2	3	4	5	6	7	8	9	10	
Rho[c]	UF	.93	.91	.97	.85	.91	.90	.98	.97	.97	.98	.950
	R(V)	.95	.87	.97	.81	.95	.95	.95	.95	1.00	.98	.950
	RS	.95	.87	.97	.81	.95	.95	.95	.95	1.00	.98	.950
PM	UF	.93	.92	.93	.84	.91	.92	.99	.96	.99	.92	.925
	R(V)	.94	.90	.93	.79	.93	.95	.85	.94	1.00	.97	.935
	RS	.95	.87	.93	.81	.94	.94	.91	.93	1.00	.92	.930
MABS	UF	9	9	11	11	10	10	5	7	3	14	9.5
	R(V)	9	7	10	14	8	10	14	9	1	6	9.0
	RS	10	12	10	12	7	12	12	11	2	14	11.5

[a] MABS, PM, and Rho refer to mean absolute deviations, product moment
correlations, and Spearman's Rho rank order correlations.

[b] UF, F(V), and RS denote the additive utility, R(V) utility, and additive
rating scale decomposition models.

[c] Because RS and R(V) are monotonically related, the necessarily have the
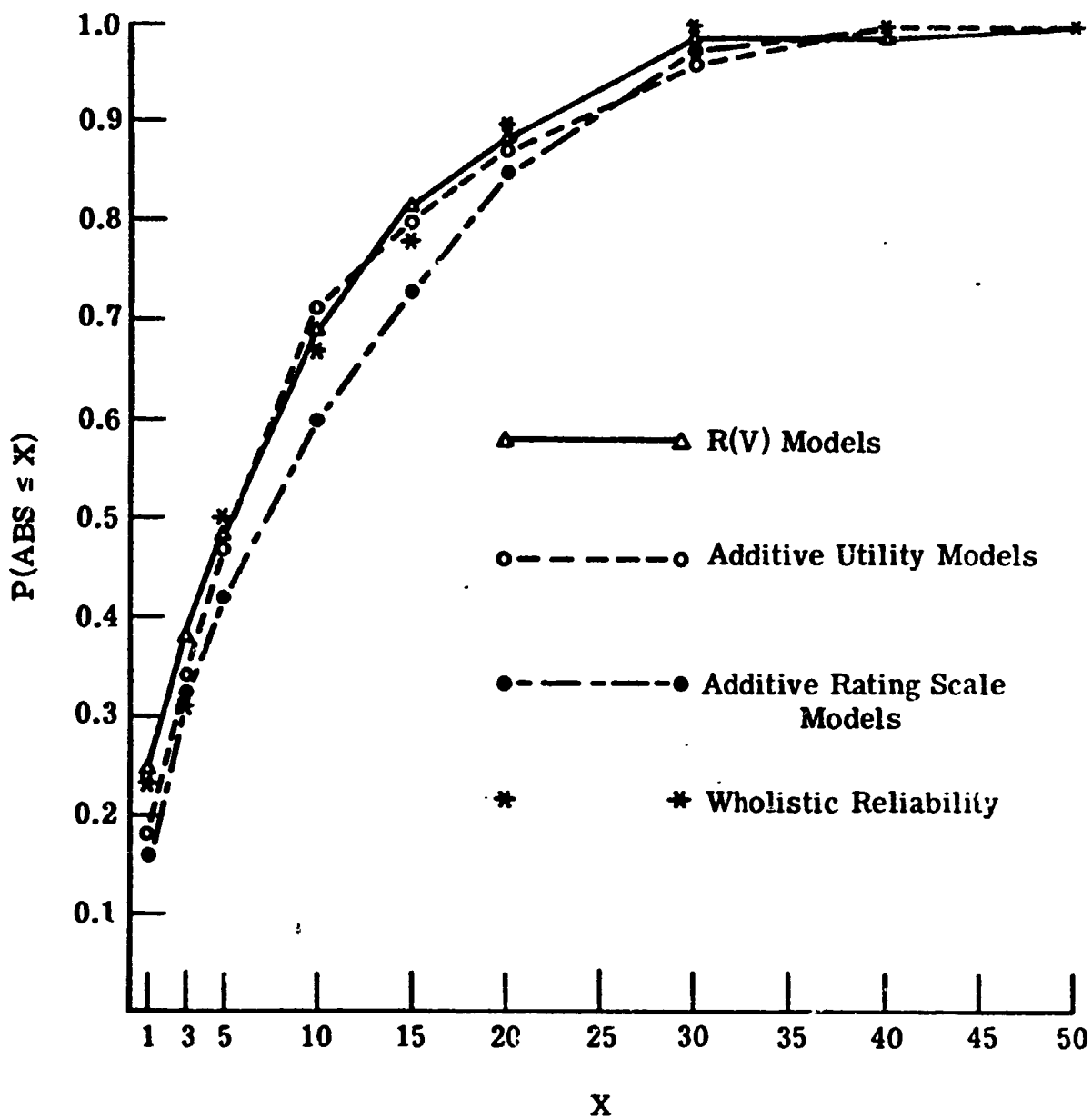same rank order correlation with the wholistic assessments.

Figure 4. Cumulative distributions of absolute
deviations from wholistic utilities.

degree of prediction afforded by the decomposition models
was about as high as possible given the degree of
reliability inherent in the wholistic utility judgments.

Convergence between Decomposition Models. As the
results presented above suggest, convergence between the
three decomposition procedures was excellent. This con-
clusion was supported both by rank order and product
moment correlations and by a MABS analysis (see Table 14).

Discussion and Conclusions

Experiment 2 provided further support for the con-
tention that additive decomposition procedures provide an
appropriate measure of riskless value. As in Experiment
1, convergence between wholistic and decomposed riskless
ratings was very high.

The second major result of Experiment 2 was somewhat
unexpected. Because the marginality assumption, which
is required for additivity under risk, is intuitively
unappealing, the experimenter had expected that subjects
would show substantial departures from additivity under
risk. The outcomes evaluated, if taken seriously, seem
to be of sufficient importance to produce violations of the
marginality assumption. Nevertheless, interaction effects

TABLE 14

CONVERGENCE BETWEEN DECOMPOSITION MODELS: EXPERIMENT 2

Measure[a]	Models[b]	1	2	3	4	5	6	7	8	9	10	Median
Rho[c]	RS-UF	.99	.92	1.00	.98	.94	.95	.97	.94	.97	1.00	.970
	RS-R(V)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
	UF-R(V)	.99	.92	1.00	.98	.94	.95	.97	.94	.97	1.00	.970
PM	RS-UF	.98	.93	.97	.98	.95	.95	.95	.93	.99	1.00	.960
	RS-R(V)	1.00	.97	1.00	.99	.99	.91	.94	.97	1.00	.96	.980
	UF-R(V)	.98	.93	.97	.98	.94	.89	.89	.91	.99	.96	.950
MABS	RS-UF	6	8	6	4	7	7	9	18	4	2	6.5
	RS-R(V)	2	8	1	6	3	22	13	4	2	12	5.0
	UF-R(V)	5	8	7	5	8	19	12	9	3	13	8.0

[a] MABS, PM, and Rho refer to mean absolute deviations, product moment correlations, and Spearman's Rho rank order correlations.

[b] UF, R(V), and RS denote the additive utility, R(V) utility, and additive rating scale decomposition models.

[c] Because R(V) is a monotonic transformation of RS, their rank order correlation is necessarily 1.00, and RS and R(V) necessarily have the same rank order correlation with UF.

accounted for but a very small proportion of the variance of the wholistic utility assessments. Thus, Experiment 2 does not support the contention that risky utility assessment requires non-additive decomposition procedures. This result should not be too strongly interpreted, however. For the outcomes evaluated were hypothetical, and consideration of the same alternatives when real jobs were at stake might have produced a different result. In addition, the possibility exists that while the job selection context does not produce non-additivity, other multi-attribute evaluation contexts will.

The third major finding of Experiment 2 was that all three decomposition procedures produced excellent, and essentially equal, prediction of the wholistic risky utility assessments. In contrast to Experiment 1, there was no evidence that the additive utility decomposition procedure was subject to a larger amount of error than the other assessment procedures. Uncritical acceptance of these results of Experiment 2 would suggest that the three decomposition methods provide equally valid measures of risky utility, and that choice between the three might be based upon simple practical considerations such as ease of assessment. That is, given the extremely high degree of convergence be ween all utility assessments, whether

wholistic or decomposed, it seems unfruitful to attempt
to determine which response mode is "best". Since they
all assign essentially the same utility to a given
alternative, it makes little difference which response
mode is adopted.

The experimenter is inclined to adopt a more cautious
approach, however. For the numerical results presented
earlier indicated that interval scale utility measures are
quite sensitive to specification of a composition rule.
When that rule should be substantially non-additive, use
of an additive rule will seriously misrepresent the decision
maker's preferences. And despite the additivity of risky
preferences in the present experimen., there may well be
real world settings where a non-additive evaluation rule
is necessary. So in applying multi-attribute utility
theory to real world problems, it would still be advisable
to employ additive assessment techniques only after having
ascertained that the decision maker's preferences are
additive. This can be accomplished by determining whether
or not the decision maker is indifferent between marginally
equivalent gambles. If he is not, then the R(V) technique
should be used. When no direct test of the additivity
assumption is feasible, the R(V) procedure should again be
used. For it is formally appropriate whether or not
preferences are additive under risk.

Finally, a brief discussion of previously unmentioned limitations of this research is in order. First, a total of only fifteen subjects participated in the two experiments. Since fairly substantial individual differences were apparent in the data, this is a serious limitation. Given the small samples used in these studies, it was impossible to explore the possibility that different assessment procedures may be required for different decision makers.

In addition, it should be noted that this research completely ignored the problem of defining the list of value attributes relevant to a given decision. These were simply given to subjects as part of their tasks. Yet in actual practice this may well be one of the most difficult aspects of an evaluation problem. It should also be noted that the practice of giving subjects a list of attributes to work with enhances the degree of convergence between wholistic and decomposed assessments. For it assures that both are based on the same information. In the real world, on the other hand, wholistic and decomposed assessments might conflict simply because the decision maker's wholistic assessments took into account considerations which he had not thought to incorporate in his decomposition model.

This criticism applies not only to the present research, but also to virtually all psychological studies of preferences for multi-attribute alternatives.

REFERENCES

Arrow, K.J. Social Choice and Individual Values, New
Haven: Yale University Press, 1952.

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

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

Coombs, C.H., Bezembinder, T.G., and Goode, F.M. Testing
expectation theories of decision making without
measuring utility or subjective probability. Journal
of Mathematical Psychology, 1967, 4, 72-103.

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

Eckenrode, R.T. Weighting multiple criteria. Management
Science, 1965, 12, #4.

Edgeworth, F.Y. Mathematical Psychics, London: C.K. Paul,
1881.

Edwards, W. Social utilities. In Decision and Risk
Analysis: Powerful New Tools for Management,
Proceedings of the Sixth Triennial Symposium, June
1971, Hoboken: The Engineering Economist, 1972,
114-124.

Fischer, G.W. Multi-dimensional value assessment for
decision making. The University of Michigan
Engineering Psychology Laboratory Technical Report
No. 037230-2-T, June 1972.

Fischer, G.W. and Peterson C.R. Ratio Versus Magnitude
Estimates of Importance Factors. The University of
Michigan Engineering Psychology Laboratory Technical
Report No. 037230-3-T, June 1972.

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

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

Hoepfl, R.T. and Huber, G.P. A study of self-explicated
utility models. Behavioral Science, 1970, 15,
408-414.

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.

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

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

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

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, 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.

O'Connor, M.F. The application of multi-attribute scaling
techniques to the development of indices of water
quality, Doctoral dissertation, The University of
Michigan, 1972.

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

Pollack, I. Action selection and 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, Reading, Mass.: Addison
Wesley, 1968.

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

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

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.

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.

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

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.
The University of Michigan, 1971.

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

Yntema, D.B. and Torgersen, W.S. Man-computer cooperation
in decisions requiring common sense. IRE Transactions
on Human Factors in Electronics, 1961, HFE-2, 20-26.