social science research institute

TECHNICAL REPORT

LEARNING AND RECOVERING ADDITIVE AND MULTIPLICATIVE
VALUE FUNCTIONS:
A CRITERION VALIDATION OF
MULTIATTRIBUTE UTILITY TECHNIQUES

Richard S. John

Detlof von Winterfeldt

December, 1982

SSRI RESEARCH REPORT 82-1

DTIC
ELECTED
JUN 3 0 1983
S D
A

83 06 30 036

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
82-1 Technical report	A129973	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
Learning and recovering additive and multiplicative value functions: A criterion validation of multiattribute utility techniques	Technical report
	6. PERFORMING ORG. REPORT NUMBER 82-1

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Richard S. John and Detlof von Winterfeldt	MDA903-81-C-0203

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS
Social Science Research Institute University of Southern California Los Angeles, Ca. 90089-1111	

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Defense Advance Research Projects Agency Department of Defense	December, 1982
	13. NUMBER OF PAGES 55

14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)	15. SECURITY CLASS. (of this report)
Cybernetics Technology Office 1400 Wilson Blvd. Arlington, Va. 22209	unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

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

18. SUPPLEMENTARY NOTES

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

validation, multiattribute utility, value, utility, structuring, decision analysis, multiple cue probability learning, simplification, functional form, additive model, multiplicative model

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

This paper describes two experimental studies in which subjects were taught additive and multiplicative value functions for the evaluation of diamonds. After learning, subjects were sent to a decision analyst who used standard multiattribute utility elicitation techniques to recover these value functions. Comparison of the taught and recovered functions allowed us to

DD FORM 1473 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

unclassified

techniques. In the real-world experiment, internal bank auditors served as subjects in a criterion validation study. Subjects provided both holistic and SMART models of commercial loan classification. Both types of models resulted, overall, in about the same level of accuracy. This level of accuracy was slightly better than a least squares solution using the same variables.

Taken together, we found the studies suggestive of two strategies for coping with complex structures in MAUM. The first is to attempt to reduce the complexity by searching for simple and independent sets of attributes that lend themselves to additive modeling. The second is to increase the model complexity, if you believe the underlying preferences are non-additive and the deviations from additivitiy are not too extreme. However, if the structures become overly complex and the deviations from additivity are too extreme, this model suggests the simple models will be prefereable to the complex ones.

Accession For

NTIS GRA&I
DTIC TAB
Unannounced
Justification

By
Distribution/

Availability Codes

Dist   Avail and/or
        Special

A

Technical Report

Learning and Recovering Additive and Multiplicative Value Functions:

A Criterion Validation of Multiattribute Utility Techniques

Richard S. John

Detlof von Winterfeldt

Social Science Research Institute

University of Southern California

Research Report 82-1

TABLE OF CONTENTS

When modeling multiattribute preferences, a decision analyst has to make three important choices:

1) choice among the basic modeling approach (e.g. riskless or risky modeling);

2) choice among aggregation rules (e.g. additive or multiplicative aggregation rules);

3) choice among elicitation techniques (e.g. trade-offs or direct rating and weighting).

The theoretical and applied literature on multiattribute value and utility assessment offers some guidance about how to make these choices. Taxonomies of decision problems can be helpful in selecting basic modeling approaches (see e.g. MacCrimmon, 1973; v. Winterfeldt, 1980; Brown and Ulvila, Note 1). Measurement theoretic independence tests can aid the analyst in identifying obviously inappropriate aggregation rules (e.g. Fishburn, 1970; Krantz, Luce, Suppes, and Tversky, 1971; Keeney and Raiffa, 1976; Dyer and Sarin, 1979). In addition, several researchers have developed criteria for evaluating the practicability and usefulness of the available elicitation techniques (see e.g. Kneppreth, Hoessel, and Johnson, Note 2; Johnson and Huber, 1977).

Nevertheless, there exists virtually no hard experimental data about the relative validity of alternative approaches, model forms, and elicitation techniques. The reason for this paucity of data is the inherent difficulty in finding a validation criterion against which to compare alternative

multiattribute utility assessments. The few existing experimental studies had to rely on convergent validation (for summaries, see Fischer, 1976; 1977; v. Winterfeldt and Fischer, 1975), experimental tests of independence assumptions (for summaries, see v. Winterfeldt, 1980), and observation of simple choices (e.g. Schoemaker and Waid, 1982). The results of these experiments indicated that, from a convergent validation point of view, the choices of a decision analyst do not matter much.

This unsatisfactory state of validation of multiattribute utility assessments grew out of the belief that utilities are simply uncheckable value statements, and that therefore no external validation criterion exists. But utilities are not necessarily uncheckable, at least not always. Decisions are made for a purpose. Often it is possible to see whether the purpose has actually been fulfilled. In addition, values do not develop in a vacuum. Rather they are learned, sometimes through explicit instructions in organizations, sometimes through outcome feedback. This offers the possibility of experimentally inducing value or utility structures in originally naive subjects and using these learned structures as a criterion for subsequent elicitation. This paradigm is closely relatd to a procedure used by Yntema and Torgerson (1961) and the multiple cue probability learning task (MCPL) task (e.g. Hammond, Stewart, Brehmer, and Steinman, 1975; Schmitt, 1978).

In a previous study (John, Edwards, and Collins, Note 3)

we used this paradigm to teach subjects value functions for the appraisal of diamonds that varied on the attributes "cut", "clarity", "color", and "carat". Subjects, who did not have preconceived notions about how diamonds should be appraised, were told that they would learn, via computer instruction, how diamonds should be evaluated. Subsequently they were presented with displays of diamonds varying on the four "C" attributes, and asked to estimate their prices. The computer then determined the "true" price through an additive value function with fixed weight ratio of 8:4:2:1. After either 60 or 120 trials subjects were able to reproduce the value function very well. (Median correlation of subjects' estimates with the true value was .93). Various elicitation methods were then applied to elicit the weights from the subjects, including formal value assessment methods, e.g. pricing out and trading-off to the most important dimension (Keeney and Raiffa, 1976); a holistic rating procedure called HOPE (Barron and Person, 1979), and direct subjective rating and ranking methods (Stillwell, Seaver, and Edwards, 1981). The direct subjective judgments of "importance" produced just as accurate weights as the formally correct assessments.

The present study goes one step beyond the question of validating alternative elicitation techniques for weighting procedures and addresses the ability of alternative techniques to determine whether a model is additive or multiplicative. An additional novel feature of this experiment was that the taught

value functions were recovered in a real life decision analysis session, in which the analyst (who did not know the taught function) had to use all the normal "tricks" of multiattribute utility assessment to test model forms and elicit value functions.

MULTIATTRIBUTE VALUE FUNCTIONS, TRADEOFF STRUCTURES, AND AGGREGATION RULES

In a measurable value model the decision maker or expert is assumed to be able to express his or her strength of preference among pairs of outcomes in the consequence space $\underline{C} \times \underline{C}$. Formally, this judgment can be represented by a quarternary relation $(a,b) \stackrel{\sim}{\scriptstyle >} (c,d)$ where $\underline{a}$, $\underline{b}$, $\underline{c}$, $\underline{d} \in \underline{C}$ and $\stackrel{\sim}{\scriptstyle >}$ is interpreted as "the strength of preference of a over b is larger than or equal to the strength of preference of c over d." Provided that certain regularity and independence conditions hold (e.g. transitivity, monotonicity), there exists a value function $\underline{v} : \underline{C} \to \mathbb{R}$ such that

$$(a,b) \stackrel{\sim}{\scriptstyle >} (c,d)$$

if and only if

$$v(a) - v(b) \geq v(c) - v(d)$$

When outcomes vary on several value relevant attributes $\underline{X}_i$, $\underline{v}$ can frequently be decomposed into some simple form, provided that independence conditions hold. Dyer and Sarin (1979) have shown that a difference value function $\underline{v}$ is multiplicative, if the order strength of preferences for

outcomes that vary only on a subset of attributes do not depend on the remaining (invariant) attributes. Furthermore, $\underline{v}$ is additive, if the strength of preference between two outcomes that vary only in one attribute is invariant under changes in the other attributes.

The resulting decompositions of $\underline{v}$ are

$$v(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} w_i v_i(x_i), \text{ or} \tag{1}$$

$$1 + Wv(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \{1 + Ww_i v_i(x_i)\}, \tag{2}$$

where

$\underline{x}_i$ is the level of outcome in attribute $\underline{X}_i$

$0 \leq \underline{v}(.) \leq 1$ is the single attribute difference value fu. ion with $\underline{v}_i(\underline{x}_{i*}) = 0$ and $\underline{v}_i(\underline{x}_i^*) = 1$ for some $\underline{x}_{i*}, \underline{x}_i^* \in \underline{X}_i$.

$0 \leq \underline{w}_i \leq 1$ is a scaling constant,

$-1 \leq \underline{W} < \infty$ is a parameter of the multiplicative model,

$\underline{v}(.)$ is the overall difference value function.

It is important to note that the additive model is a special case of the multiplicative model, in which $\underline{W} = 0$. In the following we frequently will refer to multiplicative models as including the additive form.

Multiattribute value functions

The usual procedure for obtaining $\underline{W}$ is to elicit the $\underline{w}_i$, $1 \leq \underline{i} \leq n$, and observe that Equation 2, evaluated for the best possible alternative, implies;

$$1 + W = \prod_{i=1}^{n} (1 + Ww_i), \qquad (3)$$

where $\underline{n}$ = number of attributes. Equation 3 clearly shows that $-1 \leq \underline{W} \neq 0$ will be a real root of an $(\underline{n}\text{-}1)$ degree polynomial. For 2-attributes, (3) reduces to

$$W = \frac{1 - w_1 - w_2}{w_1 \, w_2}$$

and for 3-attributes, $-1\text{-}\underline{W} \neq 0$ is the real solution of the quadratic formula, where $W = (-b \pm) \sqrt{b^2 - 4ac}/2ac$

$$a = w_1 w_2 w_3,$$

$$b = w_1 w_2 + w_2 w_3 + w_1 w_3 \text{ , and}$$

$$c = w_1 + w_2 + w_3 - 1.$$

There is no explicit solution to the problem of finding roots of a polynomial of degree 3 or more; thus, $\underline{W}$ must be determined by iterative procedures (e.g., Newton-Raphson method) for models with four or more attributes.

The usual method for assessing $\underline{w}_i$ involve $\underline{n}$ different "extreme" elements in the alternative space. The standard procedure is to elicit strengths of preference (compared to the worst possible alternative) for those alternatives whose outcome levels on each attribute are either the best possible or the worst possible. More details of elicitation will come later. The point we wish to make now, however, is that the mathematics of the multiplicative model does not impose these assessment strategies on us. Once single attribute value functions have been determined, any $\underline{n}$ strength of preference

judgments will define $\underline{n}$ equations, which combined with Equation 3 will serve to completely determine the $(\underline{n} + 1)$ scaling parameters $\underline{W}$, $\underline{w}_i (1 \leq \underline{i} \leq \underline{n})$.

In particular, scaling parameters may be specified in the two attribute case from a judgment of the strength of preference about the "middle" alternative $(\underline{v}_1 = \underline{v}_2 = .5)$, and a direct ratio judgment about the relative importance of $\underline{w}_1$ and $\underline{w}_2$. One feature of the usual assessment method is that a model consistent with Equation 2 can always be specified. As we shall see, other sets of n equations -- including those just suggested in the 2-attribute case -- may not yield a solution consistent with the model form in Equation 2.

Insert Figure 1 about here

The top two graphs in Figure 1 display plots of indifference curves for moderately substituting $(\underline{W} < 0)$ and complementing $(\underline{W} > 0)$ 2-attribute value models. The middle two plots illustrate the most extreme substituting and complementing models possible under the constraints of Equation 2.

These indifference curves are obtained by setting $\underline{v}(\cdot)$ in Equation 2 equal to a constant $(.1, .2, \ldots, .9)$, and plotting $\underline{v}_1$ vs. $\underline{v}_2$. It follows from an elementary theorem in analytical geometry that all curves of this form are hyperbolas, regardless of the sign of $\underline{W}$. Rotating the $\underline{v}_1$, $\underline{v}_2$ axis

$$(y_1 = v_1 \cos\pi/4 + v_2 \sin\pi/4$$

$$\text{and } y_2 = -v_1 \sin\pi/4 + v_2 \cos\pi/4),$$

Equation 2 ($\underline{n} = 2$) can be written in the standard hyperbolic form:

$$1 = \frac{(y_1 - h)^2}{a} - \frac{(y_2 - k)^2}{b}$$

where,

$$h = -\frac{\sqrt{2}}{2} \frac{(w_1 + w_2)}{(1 - w_1 - w_2)},$$

$$k = \frac{\sqrt{2}}{2} \frac{(w_2 - w_1)}{(1 - w_1 - w_2)}, \text{ and}$$

$$a = b = \frac{2\{((w_1 w_2)/(1 - w_1 - w_2)) - V(.)\}}{(1 - w_1 - w_2)}$$

The final two plots are indifference curves for the lexicographic disjunctive and conjunctive rules that depend only on the maximum or minimum attribute values. A disjunctive rule selects that alternative with the most outstanding quality, regardless of the other attributes, while the conjunctive rule requires that the chosen alternative satisfy minimum levels on all attributes. It should be clear that multiplicative models provide a natural and continuous bridge linking both the disjunctive and conjunctive rules, from "opposite" directions, to the additive.

There are some important features of the multiplicative rules that are not shared by additive, disjunctive, or conjunctive rules, however. First, multiplicative indifference curves are not evenly spaced, as they are for the other three.

That is, multiplicative rules are differentially sensitive to value differences, depending upon the location of the alternatives in the space. In particular, substituting rules ($\underline{W}<0$) will be more sensitive to poor alternatives (those in the lower left corner) than to excellent alternatives (in the upper right corner). Indifference curves in the upper right corner are spaced rather far apart, indicating that value differences in this region will be relatively smaller than value differences in the lower left corner, where indifference curves are more tightly packed. Small shifts in poor alternatives will result in relatively large value shifts.

This effect is exactly reversed for complementing models, where small changes in good alternatives will be easily detected by the tightly spaced indifference curves in the upper right corner. In contrast, changes in bad alternatives will be hardly detected by the widely spread indifference curves in the lower left corner.

Another peculiarity evident in only the multiplicative models is that the degree to which attributes compensate for one another varies across the space of alternatives. The trade-off relation is determined by the slope of the indifference curve in an additive model, and this slope is constant throughout the space. Disjunctive and conjunctive rules are completely noncompensating anywhere, but this holds throughout the space of alternatives, just like the additive. As demonstrated in the plots, however, the curvature of

multiplicative indifference curves varies throughout the space. Multiplicative substituting models define virtually additive trade-off relations for poor alternatives (lower left corner), while providing an almost completely disjunctive trade-off relation for excellent alternatives (upper right corner). In between these two extremes, attributes substitute to varying degrees. As before, this pattern is reversed for multiplicative complementing models. An additive trade-off provides a good approximation to the indifference curves for a complementing model in the region of good alternatives, but the curves in the region of poor alternatives approach a conjunctive rule.

Figure 1 also illustrates how differential sensitivity and commensurability are mediated by how extreme the multiplicative model is. Sensitivity and commensurability are most dependent upon location in the alternative space for the most extreme multiplicative models, i.e., $\underline{W} = -1$, and $\underline{W} \to \infty$. As $\underline{W} \to^{+}_{-} 0$, trade-off relations and sensitivity become more nearly constant throughout the alternative space.

Our discussion of the structural properties of multiplicative models has suggested that a single multiplicative model will provide widely ranging sensitivity and attribute commensurability over the alternative space. This serves to highlight the potential advantages of carefully selecting strength of preference judgments when constructing a model.

Our analysis of the differential sensitivity and commensurability of multiplicative models suggest an even more important point in regards to determining agreement between different models. In comparing additive models with different weights, it has long been known that the characteristics of the alternative space (e.g., attribute intercorrelations) will often play as important a role in determining model agreement as the identities of the models. It is clear from the plots in Figure 1 that any attempt to gauge the agreement between a multiplicative and an additive model will be highly dependent upon the region of the alternative space considered. In general, an additive model will provide a good approximation to a substituting model for "poor" alternatives, but will not correspond well in the region of "good" alternatives. The exact opposite will hold for complementing models.

EXPERIMENT I

Method

Models taught

All models were two attribute value models, with linear value functions over "Carat" and "Quality." "Quality" was explained as a composite of the three attributes "Color", "Clarity", and "Cut", expressed on a percentage scale from 0% to 100%. Carat was operationalized as the diamond weight ranging from 0.1 to 1.00.

The model forms varied in terms of the tradeoff relation between "quality" and "carat". Tradeoffs were either additive or multiplicative, and multiplicative models were either complementing or substituting. Additive models were defined with either a 4:1 or 1:1 trade-off, complementing models with either a 2:1 or 1:1 trade-off, and substituting models for only the 1:1 trade-off. This design is summarized in Figure 2, which displays exact scaling parameters and indifference curves for each of the five value model conditions.

Insert Figure 2 about here

Subjects

Twenty undergraduates (17 females, 3 males) enrolled in an introductory psychology class at the University of Southern California volunteered for the experiment. All subjects received credit toward an experiment participation requirement of the course. In addition, all subjects were informed at the beginning of the experiment that they would be paid a cash bonus between $0 and $10 for their participation. It was emphasized that the exact amount of the bonus depended upon her performance during both the learning and assessment phases of the experiment. All experiment sessions were conducted individually, and each lasted from 2 to 4 hours.

Training Procedure

All subjects were told that they were participating in a study to evaluate a "computer assisted instruction" method of teaching diamond appraisal that could one day replace the years of "on the job" training required to become an expert. They were told that the computer would first display a series of 100 "diamond profiles" consisting of information about two relevant characteristics for appraising diamonds: size and quality. It was emphasized that an oral test would follow the computer instruction, and that a cash bonus would be paid at the end of the session. Subjects were told that the best possible diamond (scoring 100% on the quality index and 1.0 carat in size) was worth $10,000, and that the worst diamond (0% quality score and .01 carat in size) was worth $10. Scores on the two dimensions for each diamond profile were independently generated from a pseudo-random uniform distribution on the unit interval. Subjects saw different sets of diamond profiles, since a different random seed was used for each subject. In all 20 "samples" of 100 "diamonds", quality and size were uncorrelated.

All diamond profiles were presented on the computer screen in the format shown in the example below:

```
QUALITY:  57%  |———+———+———+———+———+———*———+———+———+———|
              0%                                      100%

SIZE:     .45  |———+———+———+———+———*———+———+———+———+———|
              0.00                                    1.00
```
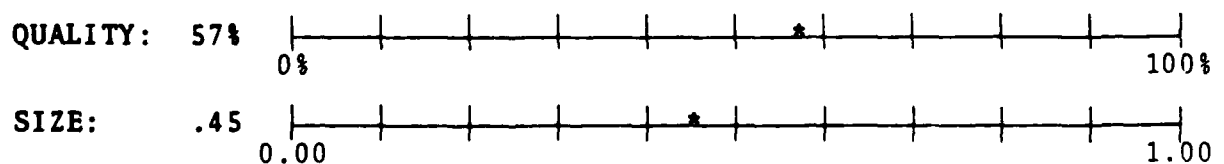
The subject then used a keyboard to type an estimate of the

worth of the diamond to the nearest $10. After checking that the estimate was between $10. amd $10,000., and requiring that the subject verify her estimate, the "true price" of the diamond was displayed, along with the amount that the subjects' estimate was over or under. When signaled by the subject to continue, the program cleared the screen and proceeded to display the next diamond profile. Time to finish all 100 learning trials varied between about 3/4 to 1 1/2 hours, as the subjects were allowed to pace themselves through the entire untimed learning task.

All "true price" feedback was computed from one of the five models defining the five "true model" conditions. (Actual prices were $10,000 times the aggregate model value, which are constrained to the unit interval.) In the case of unequal weight ratios, half of the subjects' true models assumed quality as more important than size, and half assumed the opposite. No random error was included in any model condition.

Model Assessment

Immediately following training, subjects underwent a decision analysis session designed to assess the 2-attribute value (utility) function for diamonds that had just been acquired through the 100 outcome feedback learning trials. In all cases, the analyst knew only that the true model was additive or multiplicative and that all single-attribute value functions were linear. The analyst was not even informed about the possible model parameters comprising the five true-model

conditions. One of the analysts was an expert professional who has assessed many value functions, and the other a 2nd year graduate student in psychology who has had both coursework and research experience in the area of multiattribute utility measurement.

Subjects were reminded at the outset of the elicitation sessions to consider only information about diamonds learned from the 100 feedback trials and were warned that prior notions about diamond prices would only hurt their performance (and payoff). Subjects made three types of judgments about the diamond model:

(1) value (price) differences between strategically selected diamonds and the worst (or best) diamond;

(2) ratio estimates of the relative "importance" of quality and size in determining price;

(3) judgments about outcomes (certainty equivalents) and probabilities (BRLTS) for creating indifference between a strategically chosen "sure thing" diamond, and a lottery between 2 other strategically chosen diamonds.

The order for making these judgments was randomized across subjects.

Strictly speaking only the value difference judgments are formally justified elicitation methods for recovering additive and multiplicative value functions. Ratio estimates of importance are often used as approximations for the formally

correct methods of defining scaling parameters $\underline{w}_i$ from indifference judgments (see Edwards, 1977). Lottery procedures are normally used when the evaluation has to be carried out under uncertainty. Since our training procedure did not involve any uncertainty, lottery procedures are also considered approximations. It is nevertheless of considerable interest to determine how such approximation methods fare in their relative ability to recover taught value functions.

Details of each type of assessment follows:

Value difference estimates. Subjects were asked to estimate the value of 11 strategically chosen diamonds, displayed in Figure 3. Assuming a multiplicative model with linear single-attribute value functions, scaling parameters can be determined from any two of the above judgments. For an additive model, only one is required. We obtained 11 in order to explore (a) the shapes of subjects' single-attribute value functions, and (b) the convergent validity of equivalent model parameters derived from different value difference assessments.

Insert Figure 3 about here

Subjects were instructed to make their estimates by comparing the diamonds to both the worst (worth $10.) and the best ($10,000.) possible.

Importance weights. The subject was asked to rank order the two attributes, quality and size, in terms of their

"importance" in determining price. The less important attribute was assigned a "weight" of 10, and the subject was asked to estimate the "weight" on the more important attribute, such that the ratio of the two numbers reflected the two attributes "importance ratio" in determining diamond price.

Lottery judgments. Subjects were asked to consider a series of two-outcome gambles for diamonds described in terms of their quality index and size, and various judgments about these gambles were used to (a) test additive utility independence, (b) test multiplicative utility independence, (c) assess single-attribute utility functions, and (d) assess scaling parameters for a multiplicative utility function.

Additive utility independence (AUI) was tested by asking subjects to consider two 50-50 lotteries for diamonds. Outcomes in the first lottery were either the best possible diamond (100%, 1.00) or the worst possible (0%, 0.01). The second lottery consisted of a 50-50 chance between a diamond best in quality and worst in size (100%, 0.01) and one worst in quality and best in size (0%, 1.00). This is shown graphically in Figure 4, where the first lottery results in either diamond B or W, while the second lottery results in Q or S.

Insert Figure 4 about here

Subjects not indifferent between the two lotteries were asked to indicate whether their preference was a "strong" preference

or a "weak" one. AUI requires that subjects be indifferent to the two lotteries.

Multiplicative utility independence (MUI) was tested by asking subjects to consider four 50-50 lotteries made up of the diamond pairs (1) Q and W, (2) B and S, (3) S and W, (4) B and Q. For each of the four 50-50 lotteries defined by the diamond pairs above, the analyst asked the subject to consider a sure-thing diamond "halfway" between the two. For the first lottery pair, Q=(100%, 0.01) and W=(0%, 0.01), the sure-thing diamond was defined as (50%, 0.01). The analyst then asked the subject to indicate whether she would rather play the gamble, or receive the sure-thing diamond with probability one. The quality index of the sure-thing diamond was raised up or down until the subject was indifferent between the gamble and the sure-thing. This iterative procedure was repeated for all four lotteries, each ending with the subject proclaiming indifference between the lottery and the newly specified sure-thing diamond.

MUI requires that the values for the quality index specified for the sure-thing diamonds corresponding to the first two lotteries (Q-W and B-S) be equal, as well as those specified for size corresponding to the last two lotteries (S-W and B-Q).

Single-attribute utility functions for both quality and size were assessed through a series of four outcome judgments producing indifference between a lottery and a sure-thing,

similar to those used to test MUI. The sure-thing diamond determined during MUI testing as indifferent to a 50-50 lottery between $\underline{Q}$ and $\underline{W}$ was defined for each subject as diamond $\underline{R}$. Similarly, the sure-thing determined as indifferent to the $\underline{S}$-$\underline{W}$ lottery was defined as diamond $\underline{T}$. The identity of $\underline{R}$ and $\underline{T}$ were different for each subject, and are indicated as variable in Figure 4.

The quality index of $\underline{R}$ defines one point on the single-attribute utility curve for quality. An additional point was obtained by creating a sure-thing outcome indifferent to a 50-50 lottery between diamonds $\underline{R}$ and $\underline{W}$, and a third was formed by creating a sure-thing indifferent to the $\underline{Q}$-$\underline{R}$ 50-50 lottery. Likewise, the size of diamond $\underline{T}$ defines one point on the single-attribute utility curve for size, and two additional points were obtained by creating sure-thing diamonds indifferent to a 50-50 lottery between diamonds $\underline{T}$ and $\underline{W}$, and to the $\underline{S}$-$\underline{T}$ 50-50 lottery. The determination of these sure-thing diamonds followed exactly the procedure used to construct diamonds $\underline{R}$ and $\underline{T}$ during MUI testing.

Finally, scaling parameters were assessed by asking subjects to consider a lottery between the best diamond ($\underline{B}$) and the worst ($\underline{W}$), with unspecified probabilities, and a sure-thing diamond best on quality and worst in size ($\underline{Q}$). The analyst asked the subject which she would prefer if the lottery were a 50-50 gamble between $\underline{B}$ and $\underline{W}$. The probabilities to $\underline{B}$ and $\underline{W}$ were then varied in the appropriate manner until the subject

was indifferent between the $\underline{B}$-$\underline{W}$ lottery and the sure-thing outcome, diamond $\underline{Q}$. This exact procedure was repeated with $\underline{S}$ used as the sure-thing diamond instead of $\underline{Q}$, and the lottery probabilities were again moved up or down from 50-50 to obtain indifference.

Results

Bootstrapped Regression Models

In order to verify that our subjects had actually learned a 2-attribute model for evaluating diamond worth, we bootstrapped both an additive and multiplicative model for each subject, based on the second 50 learning trial responses. The standard approach for accomplishing this is to assume the model $Y = b_0 + b\ x_1 + b\ x_2 + \epsilon$, and to estimate $\underline{b}_0$, $\underline{b}_1$, and $\underline{b}_2$ by regressing the subjects' responses, $\underline{Y}$, on diamond quality and size values, $\underline{x}_1$ and $\underline{x}_2$. However, this yields a model that is not directly comparable to either the additive or multiplicative models assumed during the subjective assessment procedures. In particular, the usual bootstrapping model allows for 3 free scaling parameters ($\underline{b}_0$, $\underline{b}_1$, $\underline{b}_2$), while the additive value (utility) model allows only one, and the multiplicative allows only two. (See equations 1 and 2 earlier). Equivalently, the standard modeling procedures assume that the value (utility) of the worst diamond (0%, 0.01) is 0.0 and that of the best diamond (100%, 1.00) is 1.0, while the bootstrapping model relaxes both of these assumptions.

The usual way of dealing with the discrepancy is to

impose a linear transformation on the bootstrapped model,

$$(Y - b_0)/(b_1 + b_2),$$

resulting in the normalizing restrictions required by the standard value and utility assessment procedures we employed. However, this method of a posteriori applying a linear transformation to a more general prediction model seems ad hoc to us. Why not just directly bootstrap a one parameter (additive) or two parameter (multiplicative) model from the subjects' responses?

One way of doing this for the additive case is to obtain the least squares estimate of a_1 in the equation

$$Y - x_2 = a_1(x_1 - x_2) + \epsilon ,$$

and write the predicted additive value, Y as

$$\hat{Y} = a_1 x_1 + (1 - a_1) x_2. \tag{3}$$

Likewise, a two parameter multiplicative model follows by obtaining least squares estimates of m_1 and m_2 in the equation

$$Y - x_1 x_2 = m_1 (x_1 - x_1 x_2) + m_2 (x_2 - x_1 x_2) + \epsilon \tag{4}$$

and writing the predicted multiplicative value (utility) as

$$\hat{Y} = m_1 x_1 + m_2 x_2 + (1 - m_1 - m_2) x_1 x_2.$$

It is important to note that regression models derived in this way will not "fit" the subjects' responses as well as those derived with more parameters. However, there is no way to predict, a priori, whether our regression models will correspond more or less to the "true" models than other

regression models with more parameters; there is no reason to expect that the true model and bootstrapped model will be closer when more parameters are estimated in the bootstrapped model.

Table 1 provides the individual results from this regression analysis and shows the close fit of the parameters of the "true" model and the model derived from regression analysis. Figure 5 is a pictorial representation of the same data in terms of indifference curves. As this Figure shows, the "true" indifference curves (dashed lines) are extremely close to the bootstrapped indifference curves (solid lines) except for one subject (No. 11).

Insert Table 1 and Figure 5 about here

In addition to these analyses, the expected correlation between values for each bootstrapped model and the true model were computed for each subject, assuming x_1 and x_2 independent, uniformly distributed on the unit interval (same as the training conditions). Mean values across the four subjects in each true model condition are presented in the top panel of Table 1. The multiplicative regression model yields expected correlations above .99 in all cases except the additive, steep weights case, which results in an expected correlation only slightly lower. Although the additive regression model is comparable for the additive model conditions, mean expected

correlations in the true multiplicative model conditions are significantly attenuated.

Insert Table 2 about here

Expected correlations are one indication of the degree of correspondence of different value models, but monotonicity tends to make correlations a rather insensitive index to model deviations. In addition, differences that do appear are often dependent upon the multivariate distribution of alternatives assumed. Moreover, correlations are relevant only if there is some reason to generate a complete ordering of possible alternatives, which is not the case in the common decision problem of choosing the one and only "best" alternative.

Thus, we chose to explore two other measures of model deviation. Mean maximum absolute differences between bootstrapped and true models are presented in the middle panel of Table 2. In terms of maximum deviations, the multiplicative bootstrapped models are much closer to the true multiplicative models than are the additive regression models. This same result is also clearly evident in the bottom panel of Table 2, in which model deviations are squared and "summed" across the entire space of possible diamonds. Regardless of the correspondence index used when a multiplicative model was taught, the multiplicative regression model of subjects' last fifty responses is significantly closer to the true model than

is the additive regression model.   Subjects did in fact learn to judge diamond worth in a non-additive manner.

Analyst Assessed Models.

To what extent were our two analysts able to recover the value models subjects learned so well? Four models were derived from the judgments subjects made following model learning: (1) a multiplicative value-difference model, (2) an additive "importance weight" model, (3) a hybrid multiplicative value model incorporating the elicited "weight" ratio, and (4) a multiplicative utility model.

Scaling parameters for the additive importance-weight model were derived to be consistent with the subjects' judgment of the weight ratio and the additivity assumption, i.e., that the two parameters sum to 1.0.

The usual multiplicative value-difference model was derived using only subjects' value-difference judgments of the two "corner" diamonds shown in Figure 3, i.e., (100%, 0.01) and (0%, 1.00).   It is easy to show that the multiplicative model requires that the two scaling parameters, w_1 and w_2, be equal to these two value-difference estimates, i.e.,

$$w_1 = (v(100\%, 0.01) - 10)/10000$$

and

$$w_2 = (v(0\%, 1.00) - 10)/10000$$

The hybrid multiplicative value model was derived to be consistent with (a) the subjects' importance-weight ratio

judgment, and (b) the value-difference judgment for the "middle" diamond, (50%, 0.50), shown in Figure 3. If $\underline{R}$ is the subject's ratio judgment of $\underline{w}_1/\underline{w}_2$, and $10000 \cdot \underline{M}$ is the subject's judgment of (50%, 0.50) - \$10, then we can solve for $\underline{w}_1$ and $\underline{w}_2$ from the equations

$$M = \hat{w}_1(.5) + \hat{w}_2(.5) + (1-\hat{w}_1-\hat{w}_1)(.5)(.5)$$

and

$$R = \hat{\underline{w}}_1/\hat{\underline{w}}_2$$

Note that when $\underline{R} = 1$, the multiplicative model requires

$$.25 \leq \underline{M} \leq .75$$

with equality holding only for $1- \underline{w}_1 -\underline{w}_2 = \pm 1$ or 1. For $\underline{R} \neq 1$, this restriction may be even more severe. Judgments of $\underline{M}$ for 6 subjects fell outside of the necessary interval; however, the hybrid model was constructed so as to be the "most extreme" possible, given the subjects' judgment of $\underline{R}$. For a substituting model, this "most extreme" solution could be determined exactly; for a complementing model, we used a solution arbitrarily close to the most extreme, since for any complementing model consistent with $\underline{R}$, there are others slightly more extreme.

Finally, a multiplicative utility model was constructed to be consistent with subjects' final two lottery judgments. It is easy to show that $\underline{w}_1$ must be the probability assigned to

B in order for the $\underline{B}$-$\underline{W}$ lottery to be equivalent to receiving diamond $\underline{Q}$ for sure. Likewise, $\underline{w}_2$ must be the probability of $\underline{B}$ making the $\underline{B}$-$\underline{W}$ lottery equivalent to a sure-thing outcome of $\underline{S}$.

Figures 6-9 are a graphical depiction of the correspondence between the indifference curves derived from the "true" model, and the indifference curves derived from the elicitation sessions. Each figure presents the results for one particular elicitation technique. Several patterns emerge from inspection of these figures.

Insert Figures 6-9 about here

First, there exists no clear cut difference between the abilities of the "expert" and "novice" analyst to match the "true" indifference curves. There appears, however, to exist some method variability, although the picture is far from clear cut.

The value difference elicitation recovered the value functions of nine subjects extremely well. In particular in the case of equal weights and multiplicative value functions, this method recovered the sign of the interaction parameter and the extent of the interaction remarkably well. Value difference elicitation appeared less well suited to pick up the value functions in the cases of unequal weights.

The additive ratio weight model did predictably poorly in the case of the multiplicative value functions and it was only

a marginal improvement on the value function elicitation in the case of additive functions. When the ratio weights were combined with multiplicative functions, the matching ability was markedly improved, in particular in the "most complicated" case (unequal weights, multiplicative value function). Utility function elicitation was a clear degradation when compared to value function elicitation, in almost all conditions. This was to be expected, as utility elicitation is, strictly speaking, not the formally correct method for eliciting value functions.

Insert Table 3 about here

The top panel of Table 3 reveals several important results about the mean values of expected correlations between each of the assessed models and the true model. First, in terms of expected correlations of outcomes, all models are quite good; 19 of the 20 mean expected correlations are in the 90's. Second, in 7 of the 8 cases, the expected correlations for equal weight additive and complementing models are higher than those for corresponding unequal weight models. Third, mean expected correlations using the additive "importance weight" model are attenuated when the true model is multiplicative. Fourth, the utility model correlations are severely depressed for the two unequal weight conditions. Finally, the value-difference and hybrid models yield highly comparable expected correlations across all five true model

conditions.

Mean maximum deviations and mean total squared deviations are presented in the middle and bottom panels, respectively, of Table 3. All of the patterns discussed above for expected correlations are born out by both of the normed distance measures. The extreme sensitivity of deviation norms, however, and especially the max norm, causes most of the differences noted in the top panel to be accentuated in the lower two panels.

It should be noted that all four of our assessed models were derived under the assumption of linear single-attribute value functions for quality and size. Many subjects gave responses that in fact indicated exactly linear single-attribute value functions. Non-linear patterns of responses to stimuli along the axes in Figures 3 and 4 were interpreted as random deviations from linearity, since none could be represented by a strictly convex or concave value function. Since there was no direct single-attribute transformation placed on quality and size in the models taught, this result is not surprising.

Structural Analysis

In addition to comparing assessed and true models via expected correlations and normed distance measures, we explored the more qualitative structural aspects of the models.

Insert Table 4 about here

The top panel of Table 4 presents the number of subjects in each of the 5 true-model conditions whose assessed value difference model was either complementing ($\underline{W} > 0$), additive ($\underline{W} = 0$), or substituting ($\underline{W} < 0$). The second and third panels produce the same analysis for the hybrid value-difference/importance-weight model and the utility model, respectively. The bottom panel of Table 3 displays the distribution of complementing vs. substituting models implied by observed violations of AUI. All subjects violated AUI, in that none were indifferent between the 50-50 $\underline{B}$-$\underline{W}$ lottery and the 50-50 $\underline{Q}$-$\underline{S}$ lottery (see Figure 4).

For the unequal weights conditions, the AUI test identified the sign of the interaction parameter equally often correctly as incorrectly. The test did, however, quite well in identifying the sign of the interaction parameter for the equal weight conditions. In all four cases in which $\underline{W} > 0$, AUI was violated by a preference for the "extreme outcomes" gamble for $\underline{W}$ and $\underline{B}$. Similarly, in all four cases for which $\underline{W} < 0$ AUI was violated by a preference for the "middle outcomes" gamble for $\underline{Q}$ and $\underline{S}$. Both preferences are consistent with the sign of $\underline{W}$. In the additive equal weights case ($\underline{W} = 0$), AUI was violated in three cases by multiattribute risk aversion, in one case by multiattribute risk proneness (see v. Winterfeldt, 1980).

These patterns indicate that a utility model is able to pick up the riskless interactions, unless the true model is truly additive.

Several other results emerge from Table 4. Overall, assessment of the correct model structure is quite good. Within the three multiplicative true-model conditions, only two of the twelve assessed value-difference models were structurally incorrect. The same result held for the hybrid value model. True additive models tended to be somewhat harder to correctly detect. In addition, both of the unequal weight true model conditions yielded more incorrect classifications than did their equal weight counterparts.

There was no clear structural shift from assessed value models to assessed utility models. In particular, the anticipated shift toward a multiattribute risk averse (substituting) utility model did not occur in any of the true-model conditions, with the possible exception of the unequal weight, complementing condition. As for value models, diagnosis of model structure via utility assessments tended to be hampered by additive trade-offs and/or unequal weights in the true model.

Discussion

Experiment I demonstrated that subjects could learn non-additive trade-off relations, and that these newly acquired value structures could be successfully discovered via standard

multiattribute value and utility assessment procedures. We found that all versions of the multiplicative value (utility) model were an improvement over the additive (importance-weight) value model when the true model was multiplicative. However, the multiplicative value model assessments were not particularly successful in detecting additivity when the true model was in fact additive. The unequal weight true-models were somewhat more difficult to assess than the equal weight models, and this was particularly evident for the assessed utility models. Non-linear single-attribute functions were not detectable.

Although the results of Experiment I are intriguing, they are not without some qualification. One of the most serious caveats is the restriction to a two-attribute stimulus domain. To what extent do our findings about learning and assessment of additive and multiplicative value functions hold in contexts involving more than 2 attributes? The purpose of Experiment II is to examine the replicability of our results in a four-attribute domain.

EXPERIMENT II

Method

Design Overview

Ten undergraduates were taught one of five different four-attribute models of diamond worth. The training procedure was similar to that in Experiment I, except that diamonds were

described in terms of the "four $\underline{C}$ s"; cut color, clarity, and carat. Just as in Experiment I, true models were either additive, complementing, or substituting. Weights for additive and complementing models were either all equal, or in the ratio 4:3:2:1. Only an equal weights substituting model was defined. Exact model parameters are given in the top portion of Table 5. Following training, value and utility model assessments analogous to those in Experiment I were performed.

Insert Table 5 about here

Subjects

Ten undergraduates (8 females, 2 males) volunteered under the same contingencies as outlined for Experiment I.

Training Procedure

The training procedure was virtually identical to that for Experiment I, except that diamond profiles consisted of four variables rather than two. An example of the 4-attribute video display is shown below:

```
CUT:     58%   |----+----+----+----+----+--*-+----+----+----+----|
               0%                                              100

COLOR:   9.5   |----+----+----+----+----+----+----+----+----+--*-|
               0                                              10.

CLARITY: 24    |----+--*-+----+----+----+----+----+----+----+----|
               0                                              100

CARAT:   .45   |----+----+----+--*-+----+----+----+----+----+----|
               0                                              1.0
```
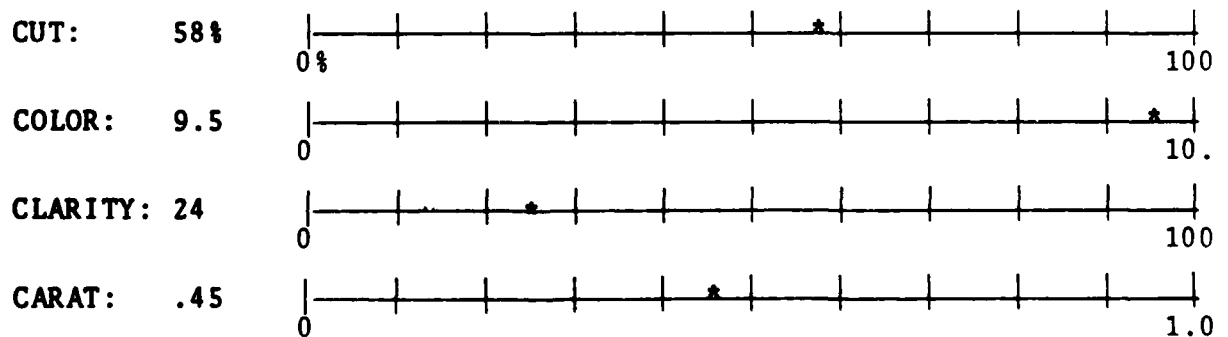
Each subject was shown a different sample of 100 diamond profiles, generated such that each of the four attributes was independently, uniformly distributed on the unit interval. (Values for color and clarity were simply multiplied by 10 and 100, respectively, for video display.) A somewhat more elaborate cover story was provided than in the first experiment, giving a rather detailed (fabricated) description of how cut, color, clarity, and carat scale scores are determined. As for Experiment I, aggregate diamond values were multiplied by $10,000, such that the worst diamond (0%, 0.0, 0.0, 0.01) was worth $10., and the best (100%, 10.0, 100, 1.00) was worth $10,000. No random error was added in any model condition.

Model Assessment.

The same two analysts from Experiment I again led subjects through an elicitation protocol immediately following training. Each analyst interacted with one subject in each of the five model conditions. At no time did analysts know what the possible model parameters were or even how many conditions there were. Analysts knew only that subjects had been taught a four-attribute multiplicative (possibly additive) model via outcome feedback.

Judgments about value-differences, "importance weights," and lotteries and sure-things were obtained in a manner analogous to Experiment I assessments. Again, the ordering for these three assessments was randomized.

Subjects estimated "importance-weight" ratios between all six pairs of attributes. Inconsistencies were simply pointed out to subjects, and a coherent set of ratios was obtained from all.

Value-difference elicitation procedures were similar to those in Experiment I. Assessments were obtained for diamonds varying in cut and color (analogous in form to those shown in Figure 5), and constant in their level of clarity ($=0$) and carat ($=.01$). One additional assessment, for cut and color both at their "best" levels, was also made. Likewise, the same 12 assessments were obtained for diamonds varying in clarity and carat, and constant in cut ($=0\%$) and color ($=0$).

Just as in Experiment I, AUI was tested by asking subjects to indicate a "weak" or "strong" preference between two lotteries (one "risky" and one "safe") with identical marginal outcome distributions. MUI was tested for all four attributes; single-attribute utility functions were elicited for all four attributes. Since MUI tests and single-attribute assessments involve only one attribute at a time, these elicitations take the same form, regardless of the number of attributes. Thus, procedures were identical to those in Experiment I. Subjects also made four BRLTS type judgments implying equivalence between a lottery resulting in either the best diamond (100%, 10, 100, 1.0) or the worst diamond (.0%, 0, 0, 0.01), and a sure thing that was best on one of the four attributes, and worst on the other three. As before,

probabilities assigned to the two extreme lottery outcomes were varied until the subject felt indifferent between the lottery and the sure thing.

Results

The only multiplicative bootstrapping model analogous to Equation 2 for 4 attributes requires the estimation of 15 free parameters. Because our assessed models use only four free scaling parameters (the fifth is determined by equation 3), a multiplicative bootstrapping model would not be comparable. Also, it is likely that a 15 parameter regression model would be susceptible to instability, due to multicollinearity among the 2, 3, and 4 way "interaction predictors," and the restricted sample size (50) of holistic responses.

Thus, we relied on correlations between the diamond price feedback and subjects' estimates over the second fifty learning trials to gauge the degree to which subjects learned the 4-attribute value model. Whereas the (expected) correlation between the true model and the bootstrapped model (used in Experiment I) is often referred to as an index of "knowledge", the correlation between actual feedback and subject responses is called "achievement." Achievement scores are virtually always smaller than knowledge scores, since disagreements due to small random inconsistencies in responding are "removed" from the knowledge index. Achievement correlations for our ten subjects are displayed in the bottom panel of Table 5. As all of the correlations are above .96, we can conclude that

subjects were able to successfully learn our 4-attribute additive and multiplicative models.

Table 6 presents two different assessments of model structures based on strength of preference judgments, and two based on lottery judgments, both as a function of the true model structure. There are two major results in Table 6. First, both of the value techniques recovered the correct structure quite accurately. Both methods were 100% correct for the 4 subjects with true complementing multiplicative trade-off structures.

Insert Table 6 about here

Secondly, there is a shift in the direction of greater attribute substitution for utility methods. One intriguing interpretation is that the model was taught under conditions of certainty, while utility models are, by design, defined over lotteries. Hence, the utility assessments may simply reflect a rather pervasive aversive attitude toward risk that is in fact meaningful, although not a part of the riskless feedback model. Another interpretation of this shift is an artifactual response-mode bias in elicitation that causes a systematic misdiagnosis of the model structures. Our data do not permit separation of these two competing hypotheses.

We also examined the agreement between assessed w_i and the true w_i. Ratios of the maximum to minimum w_i are presented

in the top panel of Table 7 for the direct "importance weight" assessments. Ratios for the value-difference and utility assessments of w_i are given in the middle and bottom panels, respectively. For the unequal w_i true model conditions, the number of rank order inversions is given in parentheses. Zero inversions indicate perfect rank order agreement; an exact opposite rank ordering would result in 6 inversions.

Insert Table 7 about here

There are two findings evident in Table 7. First, all of the assessment methods are rather poor at determining the ordinal properties of the four w_i for the multiplicative model conditions. This was true regardless of whether the model was complementing or substituting, and whether the true model w_i parameters were equal (1:1) or unequal (4:1).

Secondly, the value-difference method elicited the most extreme weights, while the utility (gamble) techniques obtained the flattest. Direct ratio judgments tended to be between these two extremes. We seem to have discovered a rather blatant response mode effect that mediates the extremeness of w_i assessments.

As was the case for Experiment I, many single attribute value and utility curves were exactly linear. Again, there were no non-linear functions that could be interpreted as srictly concave or convex, leading us to interpret deviations

from linearity as random response errors.

Discussion

We extended the finding that subjects could learn non-additive trade-off relations to the four attribute case. We also found that such complementing or substituting models could be recovered, for the most part, using the standard assessment procedures. There was a marked shift towards substitution (risk aversion) for models derived via utility theoretic methods.

None of the three assessment techniques was able to recover w_i ratios in any of the three multiplicative model conditions. There was a marked bias across all model conditions toward more extreme w_i from the value-difference assessments, and flatter w_i from the utility assessments. No strictly concave (or convex) single-attribute value or utility curves were found.

SUMMARY AND CONCLUSIONS

The most significant findings of our study were that multiplicative trade-off structures could be learned through outcome feedback and, even more importantly, that they could be reliably recovered using standard value-difference and utility assessment techniques. This result was found in the case of both 2- and 4-attribute stimuli. Assessed multiplicative models were generally better than the "importance weight"

additive model in the multiplicative true-model conditions.

In the case of four attributes, we found that ordinal information about the w_i for multiplicative model conditions was not recovered by any of the assessment techniques. Furthermore, we found that the corner point value-difference assessments produced rather extreme weight ratios, while the corner point gamble elicitations produce relatively flat ratios. There was some indication that 4-attribute utility assessments tend toward substitution (risk aversion) compared to either of the value-difference assessments of the true model.

We conclude that:

1. Multiplicative trade-off structures can be learned via outcome feedback; furthermore, these non-additive models can be recovered via standard value and utility measurement models,

2. Distinctions among value, utility, and approximate approaches are behaviorally observable, and

3. Strictly concave or convex (nonlinear) transformations of single-attribute outcome measures are not "automatically" applied before attributes are aggregated.

We will highlight some of the important questions left unanswered by these conclusions. The observed value-utility differences are important regardless of whether these are psychologically valid distinctions, or whether we are simply

exposing pervasive response mode biases. More research, perhaps with "taught" <u>utility</u> models, is necessary before these two competing interpretations can be disentangled. For now, it is clear that value and utility models derived on the same stimulus domain will not, in general, be interchangeable.

That subjects can learn and analysts can recover multiplicative trade-off structures is less open to interpretation. This clear finding suggests that if the actual trade-off structure is multiplicative, an assessed multiplicative model will perform better than an additive model. However, the question of "how much better" cannot be answered by our study. As is evident in Tables 3 and 4, conclusions about the overall level of model agreements depends upon how agreement is defined. Since agreement is obviously defined by the particular decision problem context, we cannot answer the question of "how much" in the abstract.

The three primary variables controlling model agreement that may vary from one problem context to the other are:

1. The multivariate distribution of alternatives along attributes;

2. The choice problem, e.g., choose the one best alternative, choose the best $\underline{X}$%, rank order all, etc., and

3. The standard against which difference in actual obtained value (utility) is to be compared.

Measures of agreement such as those displayed in Tables 3 and 4 make different implicit assumptions about the above three variables. For example, our expected correlation measure is predicated on a choice problem to produce an interval scaling of all alternatives, the attributes of which are mutually statistically independent.

Even if these assumptions are reasonably approximated, the practical difference beween an expected correlation of .95 and .99 will depend on the decision problem in at least two important ways: (1) What is the worst correlation obtainable, e.g., using an additive equal weights model? and (2) What are the actual value (utility) losses experienced by the .95 model relative to the .99 model? If an equal weights or random weights additive model performs at a rather low level, (e.g., below .70), then the increment from .95 to .99 seems relatively rather small. If the naive model performs at a higher level (e.g., .90+), then the relative increment from .95 to .99 takes on potentially greater significance. In addition, our perception of the benefits of a more complicated trade-off structure over a simpler one may depend on the absolute benefit of the increased accuracy. Whether the .04 increment translates into pennies or dollars depends on the decision problem in an obvious way.

Thus, the extent to which eliciting a multiplicative trade-off st ucture is justified is an open question (and probably unanswerable in the abstract). The comparability of

our standard value-difference model and the hybrid model (derived from "importance weight" judgments and one additional value difference question about the "middle" alternative) suggests an obvious practical solution for appliers of the approximate methods advocated by Edwards and others. Namely, obtain a single additional judgment of the strength of preference of the middle alternative ($\underline{v}_i$ = .5 for all $\underline{i}$) relative to the worst and best alternatives possible. Then, a multiplicative model can be derived and at least compared to the usual additive model. If this leads to an extreme multiplicative model, further assessment may be suggested.

Finally, our inability to uncover strictly concave (or convex) single-attribute function forms suggests that the standard single-attribute elicitation procedures do not suffer from obvious response mode effects that would tend to produce non-linearity when the function form is in fact linear. Closer study in which exponential single-attribute functions are taught, is warranted.

Reference Notes

1. Brown, R.V. and Ulvila, J.W. Selecting analytic approaches for decision situations. Vol. I: Overview of the methodology. (Technical Report No. TR77-7-25.) Decisions and Designs, Inc., McLean, Va., 1977.

2. Kneppreth, N.P., Hoessel, D.H., Gustafson, D.J., and Johnson, E.M. A strategy for selecting a worth assessment technique. (Technical Paper No. 280.) U.S. Army Research Institute for Behavioral and Social Sciences, Arlington, Va., 1977.

3. John, R.S., Edwards, W., & Collins, L. A comparison of importance weights for multiattribute utility analysis derived from holistic, indifference, direct subjective and rank order judgments. (SSRI Tech. Report 80-4.) Los Angeles: Univ. Of Southern California, Social Science Research Institute, 1980.

References

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

Dyer, J. and Sarin, R. Measurable multiattribute value functions. Operations Research, 1979, 27, 810-822.

Fishburn, P. Utility theory for decision making. New York: Wiley, 1970.

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

Fischer, G.W. Multidimensional utility models for risky and riskless choices. Organizational Behavior and Human Performance. 1976, 17, 127-146.

Hammond, K., Stewart, T.R., Brehmer, B.M. & Steinmann, D.D. Social Judgment Theory. In M.F. Kaplan and S. Schwartz (eds.) Human judgment and decision processes. New York: Academic press, 1975.

Johnson, E.M. and Huber, G.P. The technology for utility assessment. IEEE Transactions on Systems, Man, and Cybernetics, 1977, SMC-5, 5-15.

Keeney, R.L. & Raiffa, H.   Decisions with multiple objectives:
preferences and value tradeoffs. New York: Wiley, 1976.

Krant, D., Luce, D., Suppes, R. & Tversky, A.   Foundations of
Measurement, New York: Academic Press, 1971.

MacCrimmon, K.R.   An overview of multiple criteria decision
making. In J.L. Cochrane and M. Zeleney (eds.) Multiple
criteria decision making. Columbia, S.C.: The University
of South Carolina Press, 1973, 18-44.

Schmitt, N.   Comparison of subjective and objective weighting
techniques in changing task situations. Organizational
Behavior and Human Performance, 1978, 21, 171-188.

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

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

v. Winterfeldt, D.   Additivity and expected utility in risky
multiattribute preferences.   Journal of Mathematical
Psychology, 1980, 21, 66-82.

v. Winterfeldt, D. and Fischer, G.W.   Multiattribute utility
theory: models and assessment procedres. In D. Wendt and

C. Vlek (eds.) Utility, probability, and human decision making. Dordrecht, Holland; Reidel, 1975, p. 47-86.

Yntema, D.B. and Torgerson, W.S. Man-computer cooperation in decisions requiring common sense. IRE Transactions of the Professional Group on Human Factors in Electronics, 1961, HFE, 1, 20-26.

TABLE 1

True weights and weights of best fitting
multiplicative regression model

Subject No.	True Model			Regression Model		
	w_1	w_2	W	w_1	w_2	W
3	.50	.50	0	.48	.51	.04
4	.50	.50	0	.50	.49	.04
13	.50	.50	0	.55	.51	-.21
14	.50	.50	0	.56	.49	-.18
1	.80	.20	0	.82	.12	.61
11	.80	.20	0	.68	.57	-.64
2	.20	.80	0	.11	.89	0
12	.20	.80	0	.25	.77	-.10
7	.09	.09	99	.18	.06	70
8	.09	.09	99	.04	.22	84
17	.09	.09	99	.11	.09	85
18	.09	.09	99	.07	.14	81
9	.91	.91	-.99	.92	.74	-.97
10	.91	.91	-.99	.83	.97	-.99
19	.91	.91	-.99	.85	.88	-.98
20	.91	.91	-.99	.84	.88	-.99
5	.13	.06	99	.00	.20	∞+
15	.13	.06	99	.23	.18	14
6	.06	.13	99	.15	.12	41
16	.06	.13	99	.07	.20	52

TABLE 2

Mean Expected Correlations, Maximum
Deviations, and Total Squared Deviations
Between Bootstrapped and True Models

	True	Model	Trade-Off		

	Additive		Complementing		Substituting

Weight ratio	4:1	1:1	2:1	1:1	1:1

Bootstrapped Model

Expected Correlation

Multi-plicative	.974	.999	.990	.994	.993
Additive	.968	.999	.899	.928	.929

Maximum Deviation

Multi-plicative	.146	.034	.102	.070	.094
Additive	.112	.018	.531	.507	.506

Total Squared Deviation

Multi-plicative	.004	.000	.002	.001	.001
Additive	.004	.000	.050	.048	.048

TABLE 3

Mean Expected Correlations, Maximum Deviations, and Total Squared Deviations Between Assessed and True Models

True Model Trade-Off

	Additive		Complementing		Substituting
Weight ratio assessed model	4:1	1:1	2:1	1:1	1:1

Expected Correlation

Importance weight	.978	.969	.920	.934	.942
Value-Diff.	.945	.974	.936	.988	.983
Hybrid	.956	.974	.957	.980	.981
Utility	.856	.994	.906	.982	.964

Maximum Deviation

Importance weight	.110	.088	.529	.473	.465
Value	.363	.208	.407	.127	.180
Hybrid	.325	.206	.360	.223	.184
Utility	.450	.095	.541	.249	.323

Total Squared Deviation

Importance weight	.003	.003	.048	.048	.047
Value	.022	.015	.040	.010	.012
Hybrid	.018	.023	.030	.015	.006
Utility	.023	.003	.069	.015	.028

TABLE 4

Structural Comparisons of True Models
with Assessed Models and Utility Structure Test Results

		True Model Trade-Off				
		Complementing W > 0		Additive W = 0		Substituting W < 0
Assessment		2:1	1:1	4:1	1:1	1:1
	sgn (W)					
Value	+	2	4	3	2	0
difference	0	1	0	0	2	0
model	-	1	0	1	0	4
	+	3	3	1	2	0
Hybrid	0	0	1	1	2	0
model	-	1	0	2	0	4
	+	1	4	2	1	1
Utility	0	0	0	0	2	0
model	-	3	0	2	1	3
Additive U.	+	2	4	2	1	0
Independence	0	0	0	0	0	0
Test	-	2	0	2	3	4

TABLE 5

Experiment II: True Model Parameters
and Achievement correlations

		Trade-Off Structure				
		Additive		Complementing		Substituting
Weights		Equal	Unequal	Equal	Unequal	Equal
Para meters	W	0	0	15.000	28.230	-.750
	W	.250	.400	.067	.080	.390
	W	.250	.300	.067	.060	.390
	W	.250	.200	.067	.040	.390
	W	.250	.100	.067	.020	.390
Achievement		.987	.997	.964	.969	
		.993	.998	.971	.981	.992

TABLE 6

Structural Comparisons of True Models
with Assessed Models and AUI Tests

		Complementing W > 0	Additive W = 0	Substituting W < 0
Assessment	sgn(W)			
Value	+	4	1	1
Difference	0	0	3	0
Model	-	0	0	1
Hybrid	+	4	1	1
Model	0	0	2	0
	-	0	1	1
Utility	+	2	0	0
Model	0	0	0	0
	-	2	4	2
AUI	+	2	1	1
	0	0	0	0
test	-	2	3	1

TABLE 7

Experiment II: Assessed Weight Ratios
(Max. wt : Min. wt.)

Assessment Methods	Additive		Complementing		Substituting
			--True Maximum Wt. Ratio--		
	4:1	1:1	4:1	1:1	1:1
Direct Ratio	1:1	1.5:1	2.9:1 (3)	1.4:1	2.9:1
Import-ance	5.3:1 (0)*	1:1	3.1:1 (1)	3.5:1	4.5:1
Value Diff.	1.2:1 (0)	1:1	17.:1 (3)	1:1	2.3:1
Corner points	40:1 (0)	1:1	10.:1 (1)	5:1	10.:1
Utility Measure	1:1	1:1	3:1 (3)	1:1	1.3:1
BRLTS	2.5:1 (0)	1:1	4:1 (1)	2.5:1	1.2:1

*Note: In the case of unequal true weights, the
number of rank order inversions is given in parentheses.

Figure Captions

Figure 1.   Indifference curves for moderate multiplicative, extreme multiplicative, and lexicographic 2-attribute value functions, $V(\cdot) = .1,.2,.3,.4,.5,.6,.7,.8,.9$.

Figure 2.   Indifference curves for the five models taught in Experiment I, $V(\cdot) = .1,.2,.3,.4,.5,.6,.7,.8,.9$.

Figure 3.   Strategically chosen diamond stimuli (*) for value-difference assessment.

Figure 4.   Strategically chosen diamond stimuli (*) for utility model assessment.

Figure 5.   Indifference curves for each subject's multiplicative bootstrapped model (bold lines) and the taught model (light lines), $V(\cdot) = .25,.50,.75$.

Figure 6.   Indifference curves for each subject's elicited value-difference model (assuming linear single attribute v_i) (bold lines) and the taught model (light lines), $V(\cdot) = .25,.50,.75$.

Figure 7.   Indifference curves for each subject's elicited additive ratio "importance weight" model (assuming linear single attribute v_i) (bold lines) and the taught model (light lines), $V(\cdot) = .25,.50,.75$.

Figure 8.   Indifference curves for each subject's elicited hybrid model ratio ("importance weights" and one value difference judgment, assuming linear single attribute v_i) (bold lines) and the taught model (light lines), $V(\cdot) = .25,.50,.75$.

Figure 9.   Indifference curves for each subject's elicited

utility model (assuming linear single attribute u_i) (bold lines) and the taught model (light lines), $V(\cdot)$ = .25,.50,.75.

SUBSTITUTING (W < 0)
(W = − 8/9; w_i = .75)
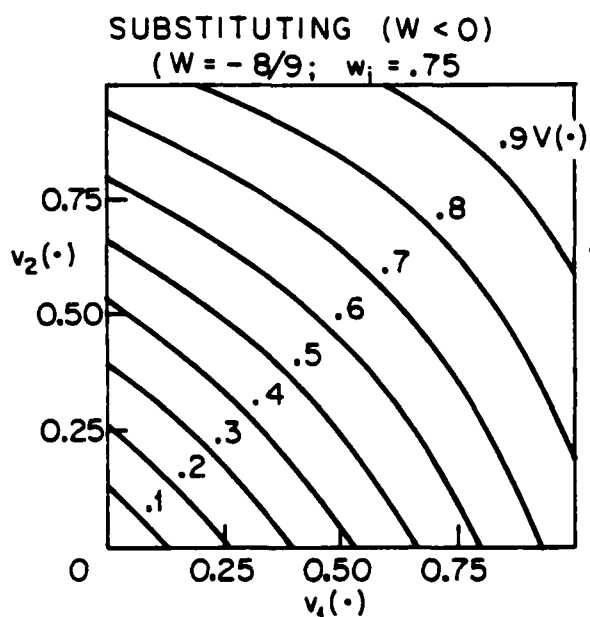
COMPLEMENTING (W > 0)
(W = 8; w_i = .25)

MODERATE MULTIPLICATIVE

$.9 V(\cdot)$
.8
.7
.6
.5
.4
.3
.2
.1

(W = − 1.; w_i = 1.0)

(W → +∞; w_i = 0.0)

MOST EXTREME MULTIPLICATIVE

DISJUNCTIVE (max rule)
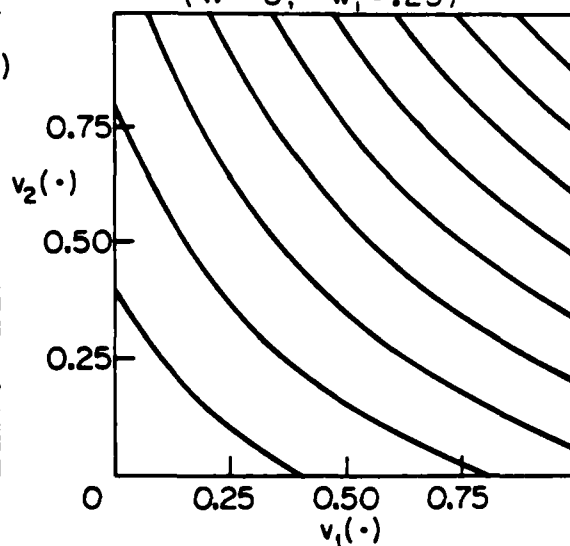
CONJUNCTIVE (min rule)
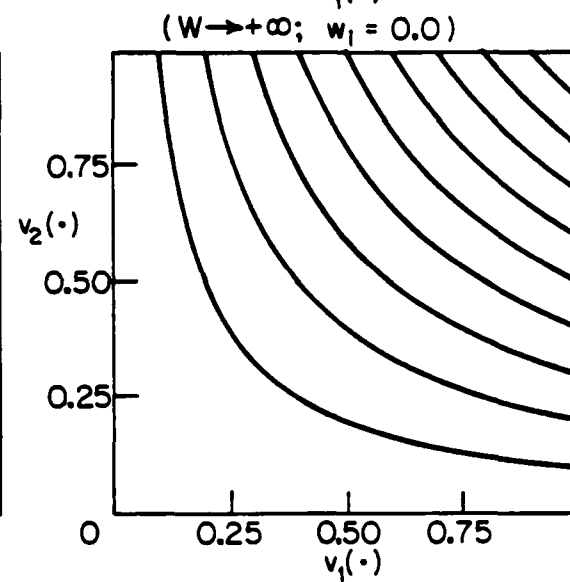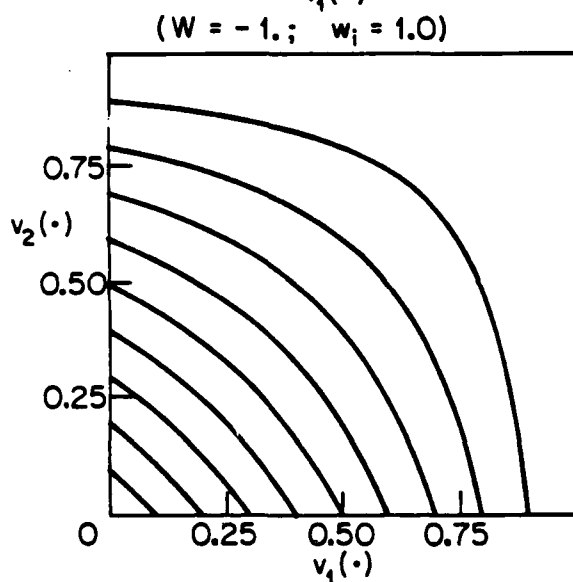
LEXICOGRAPHIC

ADDITIVE (W = O)
$w_i = .5000$

$v_2(\cdot)$

.1
.2
.3
.4
.5
.6
.7
.8
.9 = V($\cdot$)

$v_1(\cdot)$

$w_1 = w_2$
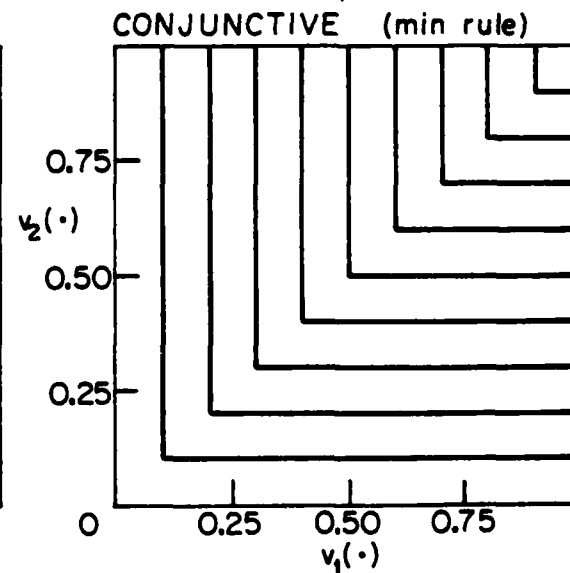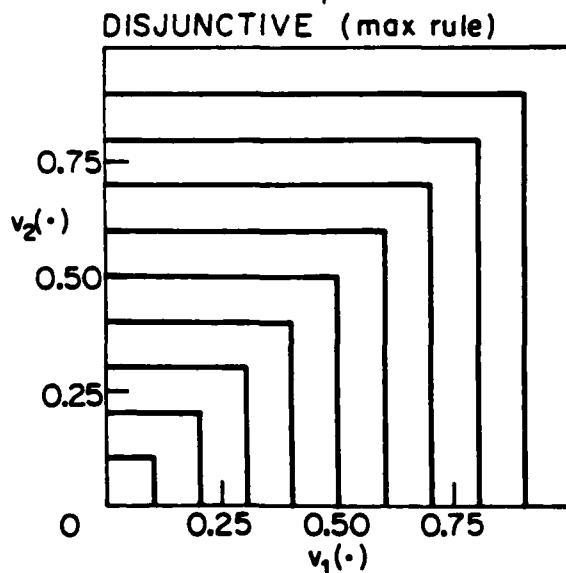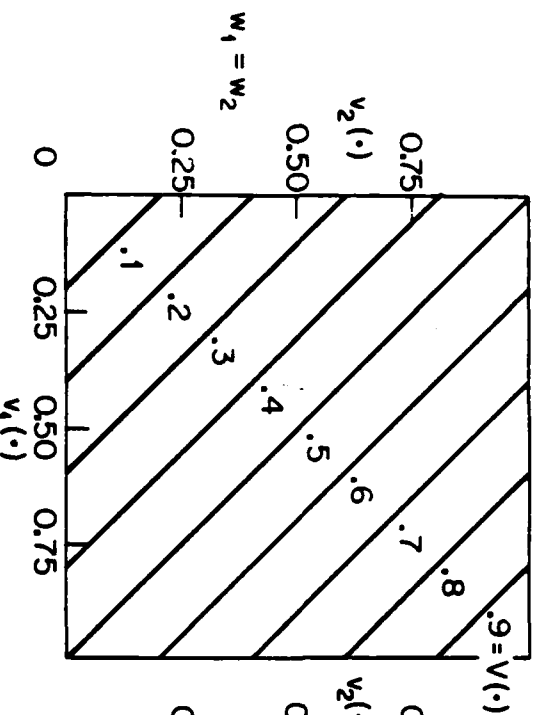
COMPLEMENTING (W = 99)
$w_i = .0909$

$v_2(\cdot)$

$v_1(\cdot)$

$w_i = .8000; \quad w_i = .2000$

$w_1 \neq w_2$

$v_2(\cdot)$

$v_1(\cdot)$
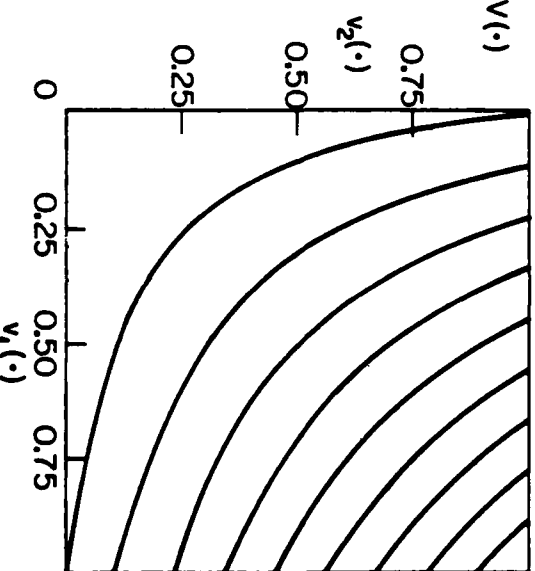
SUBSTITUTING (W = -.99)
$w_i = .9091$

$v_2(\cdot)$

$v_1(\cdot)$

$w_i = .1278; \quad w_i = .0639$

TRUE MODEL
FORM :   ADDITIVE (W=O)   ADDITIVE (W=O)   MULTIPLICATIVE COMPLEMENTING (W>O)   MULTIPLICATIVE COMPLEMENTING (W>O)   MULTIPLICATIVE SUBSTITUTING (W<O)

WEIGHT RATIO:

TRUE MODEL FORM: ADDITIVE (w=0) | ADDITIVE (w=0) | MULTIPLICATIVE COMPLEMENTING (w>0) | MULTIPLICATIVE COMPLEMENTING (w>0) | MULTIPLICATIVE SUBSTITUTING (w<0)

WEIGHT RATIO: 4:1 / 1:1 | 1:1 | 2:1 | 1:1 | 1:1

EXPERT ANALYST

NOVICE ANALYST

SIZE — 1.0 / .5 / 0

QUALITY — 0 / .5 / 1.0
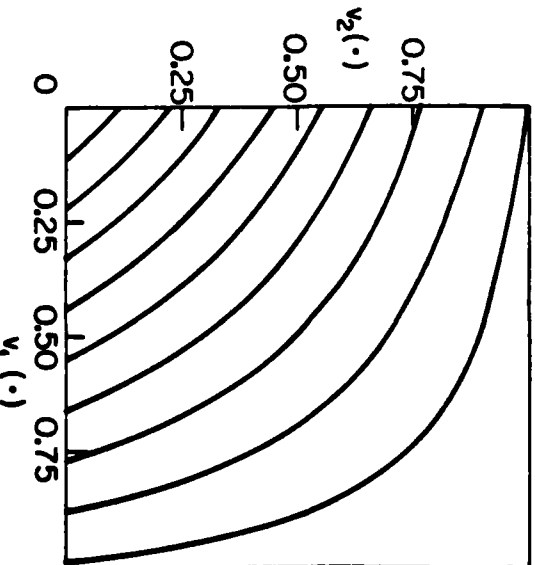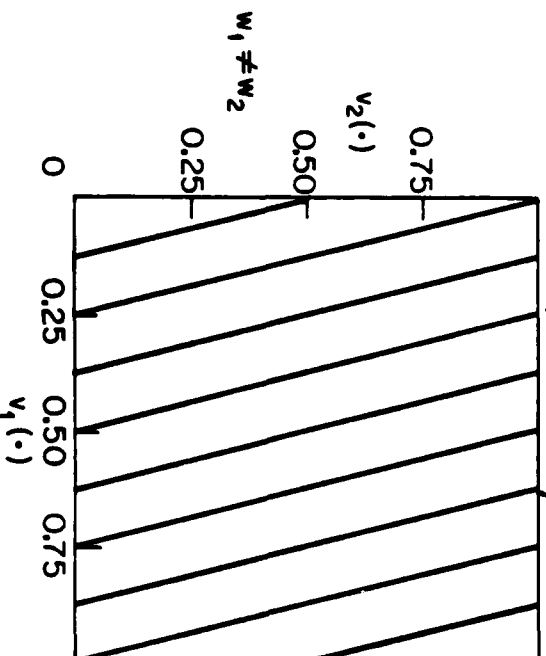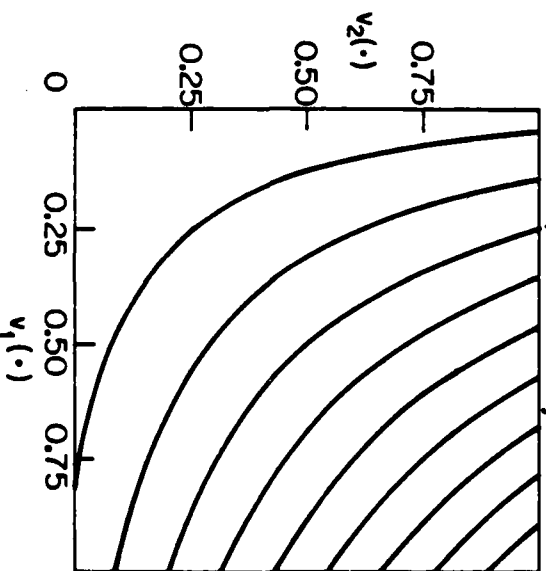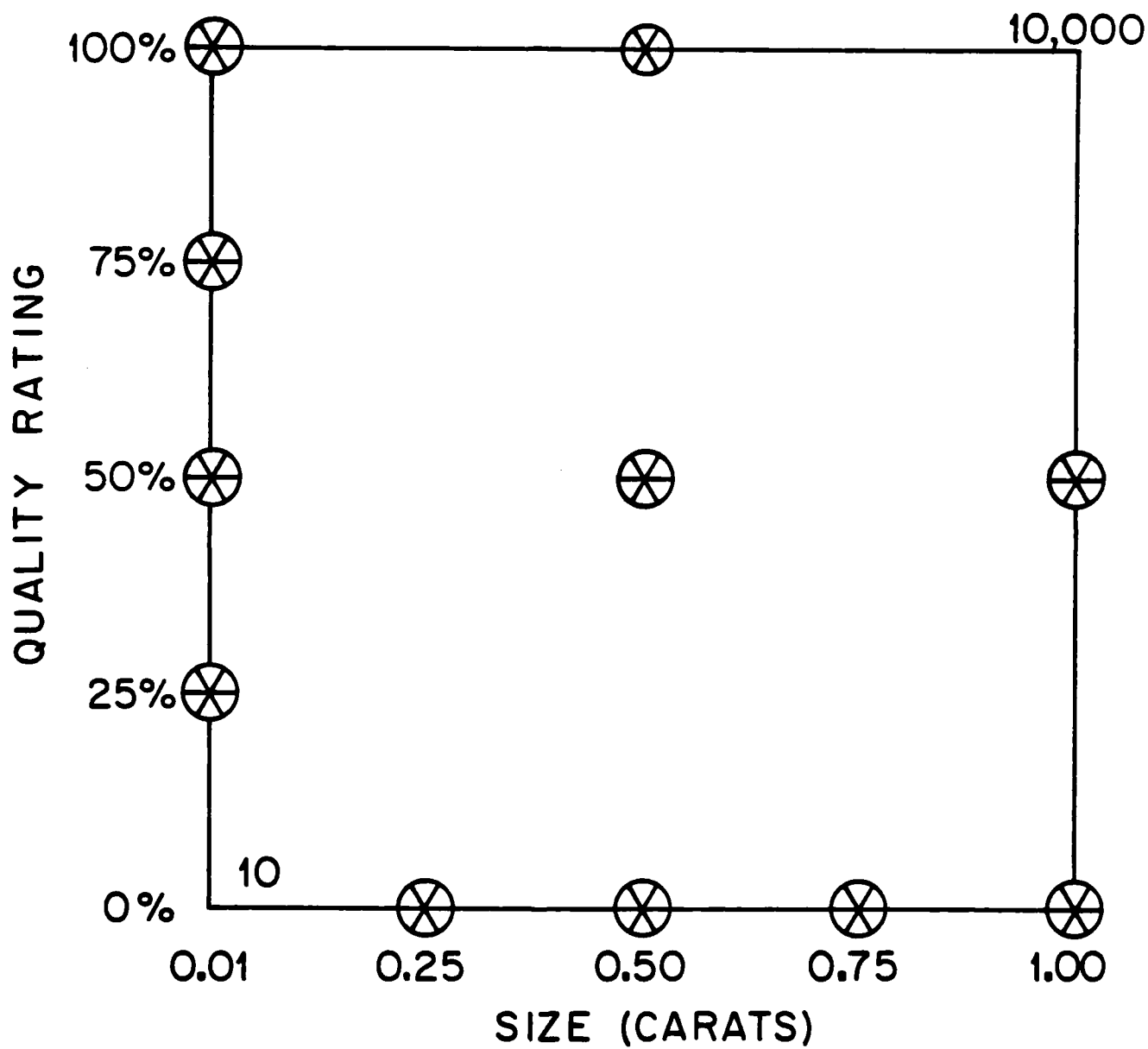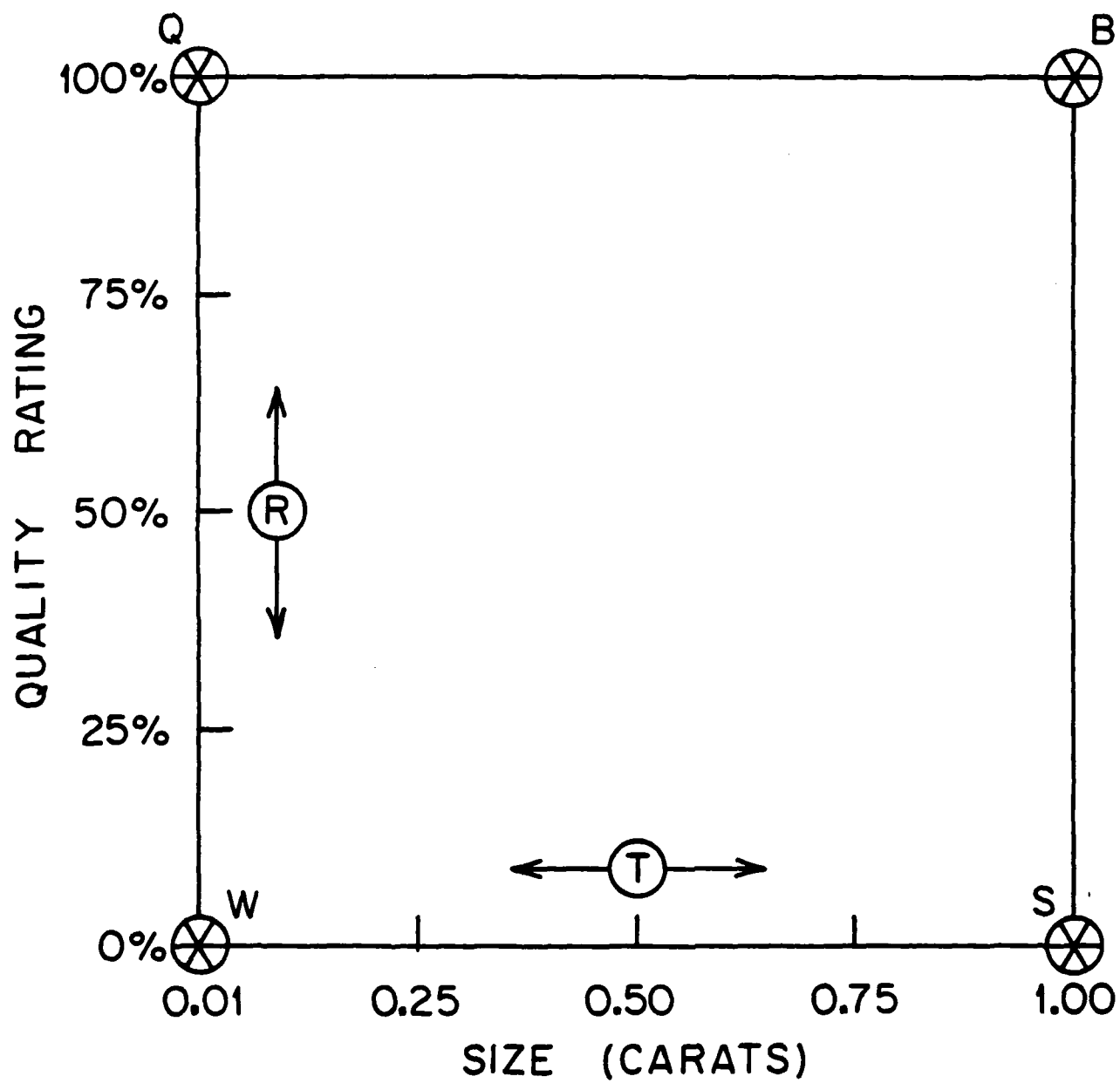
TRUE MODEL FORM:

WEIGHT RATIO:

EXPERT ANALYST

NOVICE ANALYST

ADDITIVE (W=O) 4:1
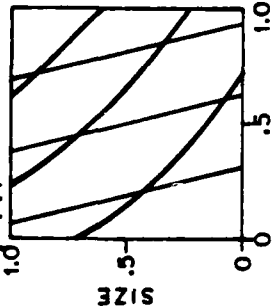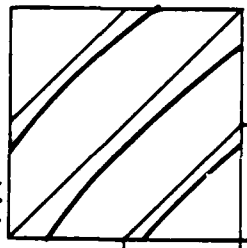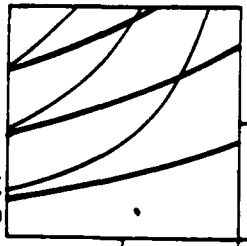
ADDITIVE (W=O) 1:1

MULTIPLICATIVE COMPLEMENTING (W>O) 2:1

MULTIPLICATIVE COMPLEMENTING (W>O) 1:1

MULTIPLICATIVE SUBSTITUTING (W<O) 1:1

QUALITY

1.0 .5 0

1.0 .5

TRUE MODEL
FORM: ADDITIVE (W=0) ADDITIVE (W=0) MULTIPLICATIVE COMPLEMENTING (W>0) MULTIPLICATIVE COMPLEMENTING (W>0) MULTIPLICATIVE SUBSTITUTING (W<0)
WEIGHT RATIO: 4:1 1:1 2:1 1:1 1:1

SIZE

1.0 .5 0

0 .5 1.0
QUALITY

EXPERT ANALYST
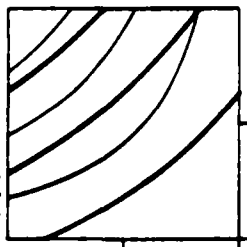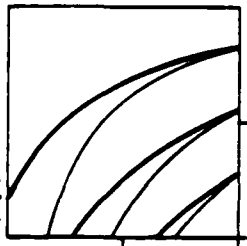
NOVICE ANALYST
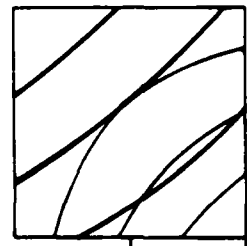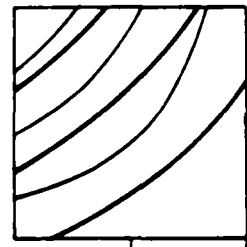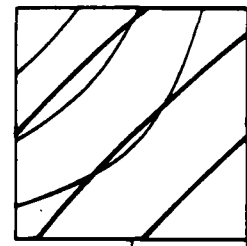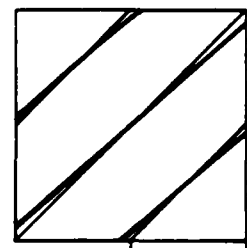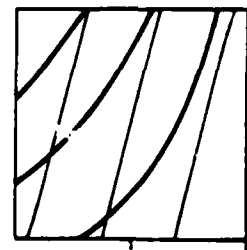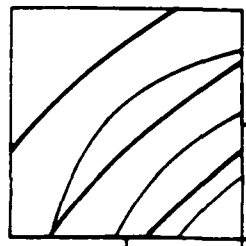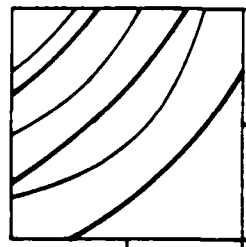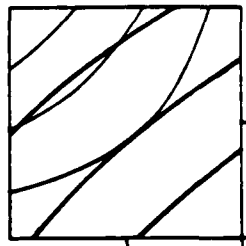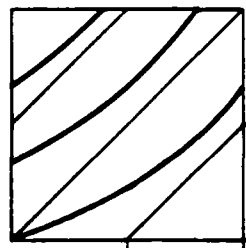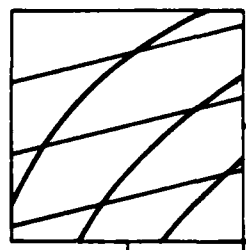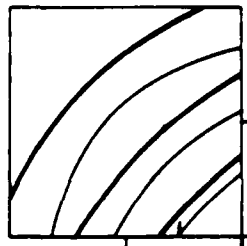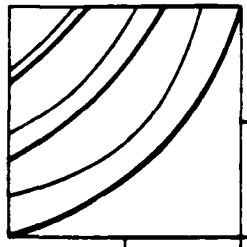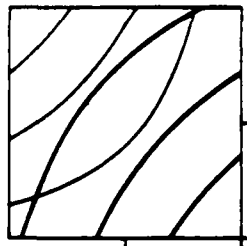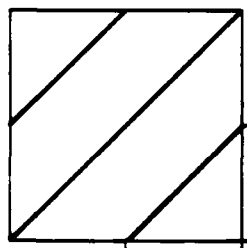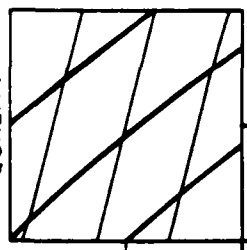
CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director   2 copies
Advanced Research Projects Agency
Attention: Program Management Office
1400 Wilson Boulevard
Arlington, Virginia 22209

Office of Naval Research   3 copies
Attention : Code 455
800 North Quincy Street
Arlington, Virginia 22217

Defense Technical Information Center   12 copies
Attention: DDC-TC
Cameron Station
Alexandria, Virgina 22314

DCASMA Baltimore Office   1 copy
Attention: Mrs. Betty L. Driskill
300 East Joppa Road
Towson, Maryland 21204

Director   6 copies
Naval Research Laboratory
Attention: Code 2627
Washington, D.C. 20375

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffins AFB
Rome, NY 13440

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, D.C. 20505

Director, Center for the Study of
  Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, D.C. 20505

Office of Life Sciences
Headquarters, National Aeronautics and
  Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, D.C. 20546

Other Institutions

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer
  Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, California 94305

Dr. Andrew P. Sage
School of Engineering and Applied
  Science
University of Virginia
Charlottesville, VA 22903

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Department of Psychology
University of Oklahoma
Attention: Dr. Charles Gettys
455 West Lindsey
Dale Hall Tower
Norman, OK 73069

Institute of Behavioral Science #3
University of Colorado
Attention: Dr. Kenneth Hammond
Room 201
Boulder, Colorado 80309

Decisions and Designs, Incorporated
Suite 600, 8400 Westpark Drive
P.O. Box 907
McLean, VA 22101

Decision Science Consortium, Inc.
Suite 421
7700 Leesburg Pike
Falls Church, VA 22043

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Naval Personnel Research and Development
   Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
   Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 54)
Attention: Dr. John Shore
Washington, D.C. 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Deputy Under Secretary of the Army
   (Operations Research)
The Pentagon, Room 2E621
Washington, D.C. 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, D.C. 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director, U.S. Army Concepts
   Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Chief, Studies and Analysis Office
Headquarters, Army Training and
   Doctrine Command
Ft. Monroe, VA 23351

Department of the Air Force

Assistant for Requirements Development
   and Acquistion Programs
Office of the Deputy Chief of Staff
   for Research and Development
The Pentagon, Room 4C331
Washington, D.C. 20330

Air Force Office of Scientific
   Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, D.C. 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and
   Operations
Directorate of Concepts (AR/XOCCC)
Attention: Major R. Linhard
The Pentagon, Room 4D 1047
Washington, D.C. 20330

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, D.C. 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense
  Research and Engineering (Test and
  Evaluation)
The Pentagon, Room 3C125
Washington, D.C. 20301

Director, Cybernetics Technology Division
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Chairman, Department of Curriculum
  Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, D.C. 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, D.C. 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, D.C. 20301

Command and Control Technical
  Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, D.C. 20301

Department of the Navy

Office of the Chief of Naval
  Operations
  (OP-951)
Washington, D.C. 20450

Office of Naval Research
Assistant Chief for Technology
  (Code: 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 No. Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 No. Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 4
800 No. Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 43
800 No. Quincy Street
Arlington, VA 22217

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
  (Code RD-1)
Washington, D.C. 20380

SOCIAL SCIENCE RESEARCH INSTITUTE
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1111
(213) 743-6955

INSTITUTE GOALS:

The goals of the Social Science Research Institute are threefold:

o To provide an environment in which scientists may pursue their
own interests in some blend of basic and methodological research
in the investigation of major social problems.

o To provide an environment in which graduate students may re-
ceive training in research theory, design and methodology
through active participation with senior researchers in ongoing
research projects.

o To disseminate information to relevant public and social
agencies in order to provide decision makers with the tools
and ideas necessary to the formulation of public social policy.

HISTORY:

The Social Science Research Institute, University of Southern California,
was established in 1972, with a staff of six. Its current staff of
researchers and support personnel numbers over 50. SSRI draws upon most
University academic Departments and Schools to make up its research
staff, e.g. Industrial and Systems Engineering, Medicine, Psychology,
Safety and Systems Management, and others. Senior researchers have
joint appointments and most actively combine research with teaching.

RESEARCH INTERESTS:

Each senior SSRI scientist is encouraged to pursue his or her own research
interests, subject to availability of funding. These interests are
diverse. Four major interests persist among groups of SSRI researchers:
crime control and criminal justice, use of administrative records for
demographic and other research purposes, exploitation of applications of
decision analysis to public decision making and program evaluation, and
evaluation of radiological procedures in medicine. But many SSRI projects
do not fall into these categories. Most projects combine the skills of
several scientists, often from different disciplines. As SSRI research
personnel change, its interests will change also.